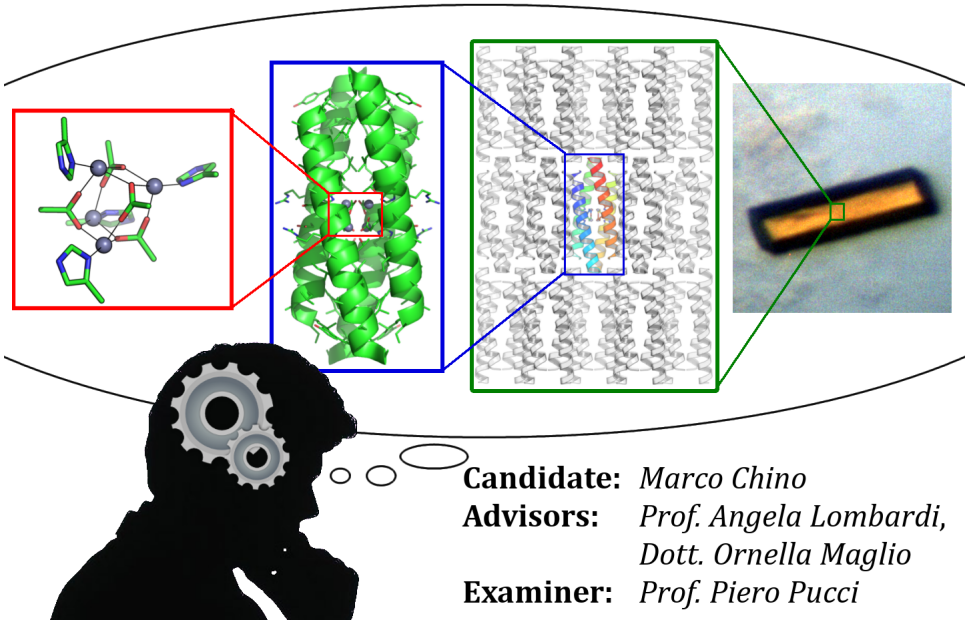
UNIVERSITY OF NAPLES

"FEDERICO II"

Ph.D. IN CHEMICAL SCIENCES - XXVI CYCLE

ENGINEERING METAL BINDING SITES IN DE NOVO DESIGNED FOUR-HELIX BUNDLES



Candidate: *Marco Chino*
Advisors: *Prof. Angela Lombardi,*
Dott. Ornella Maglio
Examiner: *Prof. Piero Pucci*

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Per nonna,

E per te.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

Table of Contents

Table of Contents

Abstract

A single polypeptide chain may access to an astronomical amount of conformers[1]. Nature selected only a trivial number of them through evolution, composing an alphabet of scaffolds, which can afford the complete set of chemical reactions needed to support life[2]. These structural templates are so stable to allow several mutations without disruption of the global folding even reaching the ability to bind several exogenous cofactors.

In this perspective metal cofactors play a crucial role in regulation and catalysis of several processes. Nature is able to modulate the chemistry of metals adopting only few ligands and slightly different geometries[3]. Thus, understanding how the fine-tuning of the scaffolds hosting them imparts the wide spectrum of reactivity is of crucial interest both in the fields of structural biology and bioinorganic chemistry.

Several scaffolds and metal binding motifs are object of intense work in the literature[4]. In this PhD work, we focused on the four-helix bundle as scaffold for metal binding sites in the context of protein *de novo* design, to obtain basic biochemical components for biosensing or catalysis.

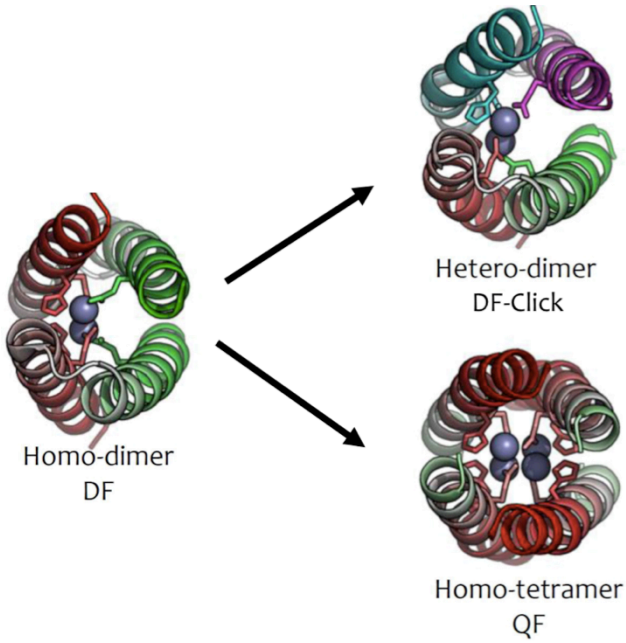


Figure 1. Top view of the DF1 dimer on the left, and the two strategies adopted in this work on the right.

In order to accomplish our objectives we chose to expand the designable space of a well-characterized *de novo* designed family of metalloproteins, the DFs (Due-Ferro). DF1, the progenitor of the family, bears a dicarboxylate bridged dinuclear metal center in its hydrophobic core constituted by the four-helix bundle unit[5]. We investigated the opportunity to find new chemical properties[6,7] by loosening on the one side and tightening on the other the C_2 symmetric environment of the metal cofactor, giving raise to two new class of *de novo* designed proteins: DF-Click and QF.

Previous attempts to design asymmetrical DFs gave successful results[8,9], even resulting in catalytic activity towards different oxidation reaction[8] upon redesign of the metal binding site[7]. We implemented an

alternative synthetic approach to generate new DF analogues by means of copper catalyzed azide-alkyne cycloaddition (CuAAC), also known as "Click Chemistry"[10]. Our design strategy led to a helix-loop-helix heterodimer, obtained through side chain chemical ligation, affording the DF-Click series.

The design, synthesis and spectroscopic characterization of an analogue of this series named DF-Click1, which was inspired to the hydroxylase component of the bacterial multicomponent monooxygenases (BMMs)[11], was completed during the PhD research activity by means of solution NMR, circular dichroism, UV-Visible absorption spectroscopy.

Further, as we were interested in the metal properties induced by an asymmetrical site, we also evaluated the opportunity to design a model with a higher degree of symmetry. We aimed to the synthesis of a D_2 symmetric homotetramer able to bind a tetranuclear metal cofactor. We named this novel *de novo* designed protein class QF (Quattro Ferro). It consists of four helix strands each one bearing the ExxH binding motif as in helix 2 of DF1. It was inspired to the cubane like Mn-O cluster found in Photosystem II, where four manganese ions provide the oxidative equivalents necessary for the four electrons oxidation of water. During the PhD work, the design process, synthesis and spectroscopic analysis of four analogues were accomplished: W45Y27 and W45Y28, which derive from the *ab initio* design of a protein crystal with predetermined topology; QF6 and QF33, whose design derives from the serendipitous finding of an already deposited crystal structure of a GCN4 mutant with a similar

crystal topology, but without the desired polarity. A combination of NMR and X-Ray analysis confirms the correctness of the design. Spectroscopic analysis intended to ascertain the metal binding stoichiometry and geometry is under course.

References

(1) Dill, K. A.; Chan, H. S. *Nat. Struct. Mol. Biol.* **1997**, *4*, 10.

(2) Sillitoe, I.; Cuff, A. L.; Dessailly, B. H.; Dawson, N. L.; Furnham, N.; Lee, D.; Lees, J. G.; Lewis, T. E.; Studer, R. A.; Rentzsch, R.; Yeats, C.; Thornton, J. M.; Orengo, C. A. *Nucleic Acids Res.* **2012**, *41*, D490.

(3) Bertini, I. *Biological Inorganic Chemistry: Structure and Reactivity*; University Science Books, 2007.

(4) Lu, Y.; Berry, S. M.; Pfister, T. D. *Chem. Rev.* **2001**, *101*, 3047.

(5) Maglio, O.; Nastri, F.; Martin de Rosales, R. T.; Faiella, M.; Pavone, V.; DeGrado, W. F.; Lombardi, A. *Comptes Rendus Chim.* **2007**, *10*, 703.

(6) Faiella, M.; Andreozzi, C.; de Rosales, R. T. M.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A. *Nat Chem Biol* **2009**, *5*, 882.

(7) Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; Degrado, W. F. *Nat. Chem.* **2012**, *4*, 900.

(8) Kaplan, J.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11566.

(9) Calhoun, J. R.; Kono, H.; Lahr, S.; Wang, W.; DeGrado, W. F.; Saven, J. G. *J. Mol. Biol.* **2003**, *334*, 1101.

(10) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.

(11) Tinberg, C. E.; Lippard, S. J. *Acc. Chem. Res.* **2011**, *44*, 280

Chapter 1: Introduction

protein folding, structural motifs, metal binding motifs, four-helix bundle, designability, de novo design, monooxygenase, Due Ferro

1. Introduction: the folding space

Conformational degeneracy of peptides give raise to an astronomical number of states which could be equally populated or not, depending on the interactions taking place among residue main chains, side chains and water, in a subtle equilibrium leading to a well-folded structure[1].

The number of possible folds, in principle, is only constrained by the forces guiding the organization of complex matter, simple physical principles, such as volume exclusion and electrostatic repulsion[2]. Nevertheless this number would be higher than the number of folds actually found in nature.

Classification projects of fold space, such as SCOP (structural classification of proteins) and CATH (class, architecture, topology, homology), undoubtedly confirm the idea of a restricted repertoire of folds[3]. There is a general agreement that the number is bounded, and it has been estimated between 1'000 and 10'000, depending on how the folds are clustered and identified. Despite the one-fold variability of this estimation, this number is incredibly small if compared to what we can expect[4]. How nature selected only a trivial number of them through evolution, composing an alphabet of scaffolds, which can

afford the complete set of chemical reactions needed to support life, is still an open question.

We may have a hint, if we define the two spaces by which each fold is defined: the sequence space and the structure space.

Sequence space shows as many dimensions as the length of the protein, each dimension enumerated by the 20 different naturally occurring amino acids. In this space, each sequence of 50 residues is a point in a 50-dimension space. Given this length, we will have a total of 20^{50} points, and we should imagine that similar sequences, based on the physicochemical features of residues in the same position, are close to each other.

Structure space is described by the combination of the coordinates of each atom composing the protein chain. Thus, for a peptide of 50 residues, considering a simple poly-alanine, we would have 3000 dimensions ($100 \times 10 \times 3$). Any possible conformation of the peptide chain would be a point in this space. We can visualize how native conformations (e.g. α-helix in the poly-Ala) would cover only a tiny fraction of this space, while normal modes of vibration of this native folds would represent a cloud of points sharing almost the same coordinates. In such a space, homologous proteins would share almost the same place (neglecting contributes arising from non-identical residues).

Given the high dimensionality of these spaces, we cannot visualize them, but we measure the distances between points in terms of sequence identity (sequence space) and root mean square deviation

(RMSD, structure space). We can go further, trying to visualize cross-correlation between the two spaces in terms of fitness of all structures with respect to the set of possible folds a protein may show. In this visualization (Figure 1), each sequence lies on a surface whose dimensions in the *xy* plane represent a feasible fold, and its third dimension is inversely proportional to the number of sequences assuming that particular fold. Thus this surface has many holes in it, since many sequences cannot be accommodated in any stable protein structure and so have no measurable fitness.

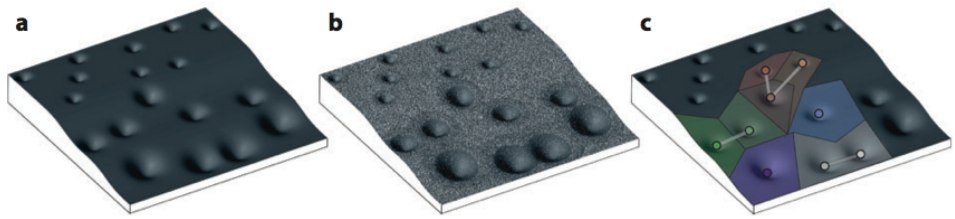


Figure 1. A schematic representation of the sequence and structure space, as function of the fold fitness. a) the smooth surface plots the fitness of all the possible sequences against all the feasible folds. The deeper is the well the bigger is the fitness. Larger wells may host several, even different sequences. b) Many sequences cannot be accommodated in any folded state, leading to the presence of several holes in this surface. These folds are not populated at all. c) We can evaluate folds based on their mutual sequence-structure relationships, evolutionary patterns may be evident in same cases, where sequence space exploration lead to structures with high fitness, thus decreasing the evolutionary pressure. Adapted from reference [3].

This visualization of the correlation between sequence and structure spaces depicts a scenario in which only few folds are robust toward sequence mutations (larger wells) and that these folds are particularly stable as they fulfill several structure requirements for correct folding (deeper wells). In the perspective of evolution, several folds may access to several functions, while exploring the sequence space for those folds, some of them were more "functional" than

others, thus leading to the selection of a discrete number of elected folds.

The recurrence of folds such as the triosephosphate isomerase (TIM) barrel[5], 3-helix bundles[6], 4-helix bundles[7], and immunoglobulin (Ig) motif[8] in proteins with highly divergent sequences undoubtedly confirms the general tendency of nature to select those scaffolds which are more prone to mutation and new functions introduction, without perturbing the global 3-dimensional structure[9].

Secondary and tertiary structural motifs are even more abundant, as they can be adopted in several structural contexts: capping motifs such as Schellman/Rose (α_L-β)-helix[10], hydrophobic staple (β-β)-helix[11], β-α_L-β loop[12], turns in β-hairpins[13].

Seminal works on the recognition of secondary and tertiary motifs had a key role in the comprehension of the mechanisms underlying in the stabilization of tertiary and quaternary folds[14]. In the era of "Structural Genomics"[15], the easy access to a great amount of structural data moves the seeking of novel motifs to a statistical level. Structural motifs are now under thorough analysis in the literature, assuming the profile of the more general concept of structural contexts. Computational algorithms can easily identify such structural contexts by means of sequence or geometric evaluations.

MSDMotif search engine[16] for secondary structure motifs is particularly powerful in the discovery or analysis of sequence patterns able to bind small ligands and cofactors. More recently DeepView, a Swiss-PDBViewer binding[17], allows more complicated structural motif

search based on tridimensional orientation even for non-contiguous residues. By means of such structural search engines, Johansson et al.[18] showed that such non-contiguous recurrent structural motifs (RSMs) can cover up to the 60% of the structure of an unknown function protein (PDB ID: 2in5), which has a very low sequence identity in the whole Protein Data Bank (PDB). Almost the full set of contacts for this protein, already categorized as New Fold in the SCOP classification, can be covered by RSMs. Computational alanine scanning of these RSMs further confirmed the role of these structural contexts in stabilizing the global fold.

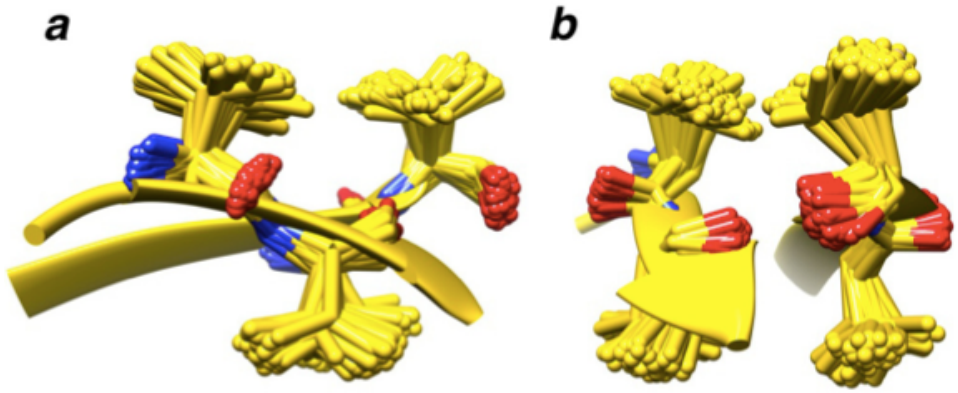


Figure 2. Orthogonal views of the most recurrent RSM, recurring in 108 mutually unrelated chains. Four leucines interact in an up-down topology, where two leucines are eclipsed in front of each other. Adapted from reference [18].

Gonzalez et al. (*unpublished results*) implemented the suns-search engine (http://degradolab.org/suns/), which makes an even faster search for structural contexts on the whole PDB, visualizing already superimposed results in the PyMOL[19] molecular graphics software.

Given the potential of such retrostructural analysis in the comprehension of the mechanisms involved in protein folding, it would be very interesting to understand the role of the metals in these secondary and tertiary structural motifs, how they are able to bind them and/or how they stabilize them. A complete analysis from this perspective is absent in the literature, but some metal binding structural motifs have found a great widespread and they will be briefly reviewed three of them: Calcium binding EF hand[20], Zinc finger binding motif[21], ExxH four-helix bundle binding motif[22].

1.1. Metal binding structural motifs

Approximately 40% of the structures deposited in the PDB contain a metal ion, including weakly bound metals of first and second group. The identity and oxidation states of the metals bound to proteins in living systems are wide-ranging. Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}, Zn^{2+}, Mn^{2+}, Ni^{2+}, $Cu^{+/2+}$, $Fe^{2+/3+}$, and $Co^{2+/3+}$ are most frequently found to bind proteins under physiological conditions[23]. Various donors are involved in the coordination of these metal ions, mainly nitrogen, oxygen, and sulfur, which are provided by amino acid side chains. Metal ions serve a variety of functions in proteins, spanning from protein structure stabilization to catalysis, signal transduction, nitrogen fixation, photosynthesis, and respiration[24].

Analysis of the metal coordination, their catalytic activity and rich redox chemistry was object of several papers, reviews and books in the field of bioinorganic chemistry[25]. This paragraph focuses on the

structural aspects of the secondary and tertiary motifs that bind metal ions, and on the general rules that can be derived.

In particular, attention is devoted to three structural motifs, which have found a great impact in the literature: calcium binding EF hand[20], zinc finger binding motif[21], ExxH four-helix bundle binding motif[22].

The classical calcium binding site in the EF hand motif consists of a helix-loop-helix, the loop consisting of 12 contiguous residues that provide oxygen atoms for metal liganding, as found in the regulatory protein calmodulin (CaM)[26]. The calcium binding site in CaM involves seven oxygen atoms at the vertices of a pentagonal bipyramid. Only five of the 12 residues in the loop provide oxygen ligands to the calcium ions, while water molecules fill the remaining coordination sites. The loop involved in metal binding was also called DxDxDG-*like* motif[16], indicating the conserved amino acidic sequence in calmodulin from different organisms. The third aspartate can be commonly mutated to an asparagine or a serine, and the first part of the loop has been also identified as an Asx-turn motif[16]. The peculiarity of this binding motif is that it is found in several non-related proteins with very different folds, such as integrin alpha N-terminal domain and tk-subtilisin (Figure 3). This complex loop, as defined by the primary, secondary and tertiary patterns, make this calcium binding motif a prototype for the concept of metal binding motifs, which plays a key role in the structure and function of several metalloproteins.

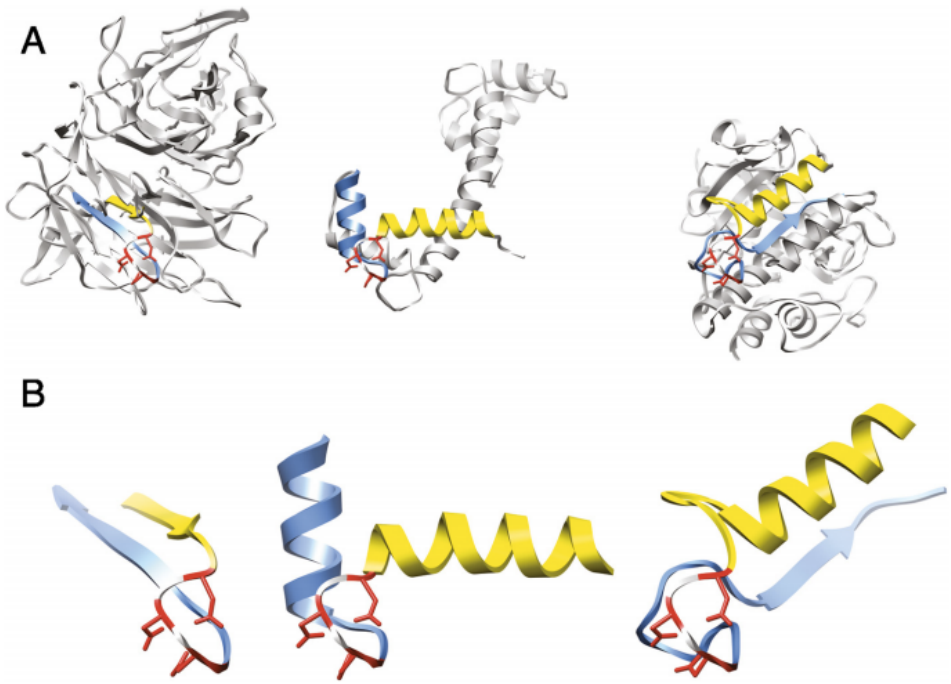


Figure 3. A) Structures of two unrelated proteins bearing the calmodulin (centre) calcium binding motif. B) Integrin alpha N-terminal domain enclose the motif in a beta fold, while subtilisin in an alpha-beta fold. Adapted from reference [17].

Zinc fingers are regions of proteins containing four residues of histidine and/or cysteine, which bind zinc in a tetrahedral geometry[27]. In this binding motif, coordinating residues form a loop that is frequently involved in the interaction with nucleic acid (finger)[28]. The classical zinc-finger bears two cysteines at a turn connecting β strands, and two histidines in an α helix. These secondary structure elements pack together in a simple β–β–α fold. Cysteines in the β-turn linking the β-strands make a full set of hydrogen bonds with the main chain stabilizing the global fold (Figure 4). Zinc finger motif does not fold in the absence of Zn(II)[9], demonstrating the structural role of metals, and

how protein evolved to fold around them. Zinc fingers are also a key example of how secondary coordination shell interactions are crucial in the affinity towards the metal and in the specificity of the fold.

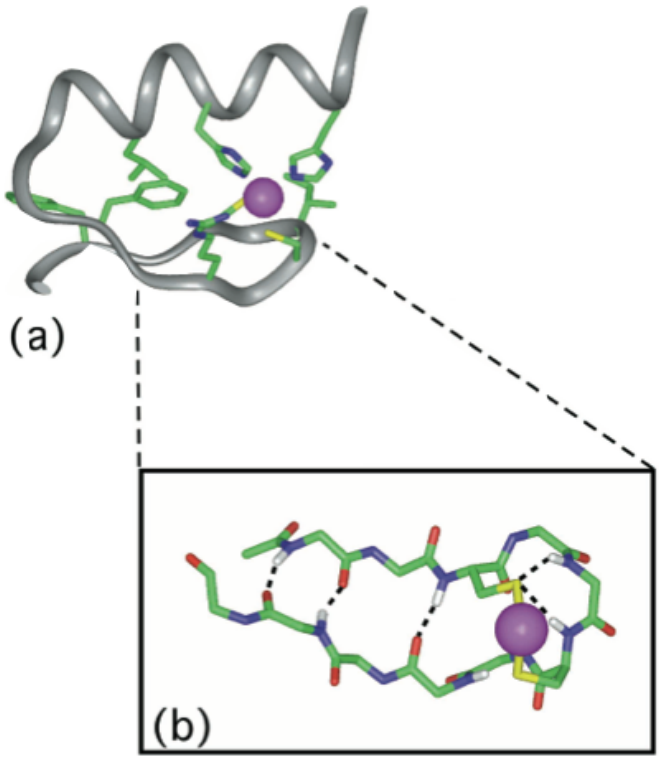


Figure 4. Zinc finger domain from ZIF268 protein[28], showing the global β–β–α global fold (a), and its b hairpin presenting the H-bond pattern crucial for folding. Adapted from reference [9].

While previous metal binding motifs have a key role in the stabilization of the global fold, ExxH motif accomplishes its metal binding function co-aided by the four-helix bundle in which it is inserted. This binding motif is central in the present work and therefore will be described in more details in the subsequent paragraphs. First, the four-helix bundle will be presented in the

contests of protein folding and design (inverse protein folding), followed by a focus of the ExxH metal binding motif in natural and *de novo* designed proteins.

1.2. Four-helix bundles and the designability concept

The four-helix bundle motif is very common in nature[7]. It is classically viewed as an α-helical coiled coil[29]. Coiled coils are characterized by a 7-residue (heptad) repeat. Residues composing the repeat are classically alphabetically ordered from *a* to *g*, with *a* and *d* positions always directed towards the interior of the bundle. Even though hydrophobic interaction by *a* and *d* positions are essential to direct the association of the helices, they don't recapitulate the full pattern of interactions needed to drive the correct folding of coiled coils.

In the special case of 4-helix bundle, four complementary-packed amphiphilic helices coil together, segregating the hydrophobic core from the solvent, thus probably constituting the easiest example of a globular protein. Helices are typically characterized by a 20° crossing angle, producing an almost ubiquitous left-handed superhelical twist[9]. Pioneer work of Harbury et al[30,31] on GCN4 mutants demonstrated how precise tuning of the side chain packing in the interior of the protein tune the properties of the bundle. He showed that leucine at position *a* and isoleucine at position *d* drive the tetramerization, thanks to the complementarity of these side chains through the bundle. However, such a recursive pattern may lead to a mixed orientation

between parallel or anti-parallel states[29]. As evidenced by the work of Yadav et al[29] and Deng et al[32], *e* and *g* positions are crucial in the relative ratio between the two orientations. Polar side chains selectively direct the orientation and the homo/hetero-tetramerization, even by destabilizing the alternative orientations[33]. While polar interactions do not influence the packing of the hydrophobic interior, hydrophobic substitutions at g positions lead only to anti-parallel tetramers with different amino acid mutual disposition, and inter-digitated helix-helix interfaces following a knobs-into-holes pattern. Limited to the anti-parallel case, three preferential topologies arise from the mutual orientation of the helices composing the four-helix bundle: the classical *a-d* core as found in ROP[34], the *a-d-e* core as found in the Lac repressor protein[35], and the *a-d-g* core as found in the hydrophobic *g* mutants of GCN4[32](Figure 5).

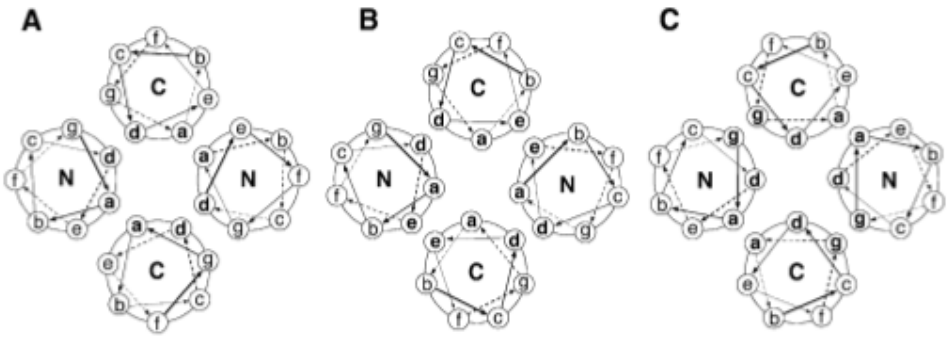


Figure 5. The three main core packing in tetrameric coiled coils. A, B, C show the three alternative core packing *a-d, a-d-e, a-d-g* respectively. Adapted from reference [32].

Such a deep understanding of the factors involved in the folding of the four-helix bundle inspired several authors to design new artificial

proteins adopting this scaffold, in a practice that is commonly identified as *de novo* design[9].

De novo design in its purest meaning consists in the ability to design a protein from scratch, planning all the secondary and tertiary structural features founding only on the first principles. This approach has been also referred to as inverse folding problem[36], since the designer wants to start with the idea of a folded polypeptide chain, possessing a determined, probably non-natural, function, and finish with a sequence, which encodes the desired information for correct structure/function. Several outstanding papers appeared in the literature describing the successful design of proteins with several global folds[37], reaching the first design of a protein with an unprecedented fold not yet found in nature, Top7[38]. A complete review of the successful examples of *de novo* design is far from the target of this thesis; it just outlines the main problems behind the design concept, and which kind of folds to prefer as targets.

As we showed before with the energy landscape theory of folding, we hypothesize that native states are sufficiently low in energy to avoid becoming trapped in local misfolded energy minima[36]. Nature, by means of evolution, explored the sequence space and selected the folds fulfilling this requirement. *De novo* protein design points to achieve minimal frustration of the selected fold via careful identification of sequences and their properties[39]. The exponential number of possible sequences determines the complexity of this search, since 20^N sequences are possible for a N-residue protein, accounting only for the

20 natural amino acids. The designer may account on several computational methods that help to narrow the search. Deterministic methods, such as the dead-end elimination (DEE)[40], self-consistent mean field[41] or site-specific probabilistic methods (e.g. SCADS[42]), help to select a set of sequences with a higher confidence to give the desired conformation.

DEE is an exhaustive search for the global minimum in a set of sequence-rotamer states[40]. It is based on the DEE theorem: For two rotamers i_r and i_t at position i, i_r is incompatible with the global minimum and will be eliminated when

$$E(i_r) + \sum_{j \neq i}^{N} \min_{s} E(i_r, j_s) > E(i_t) + \sum_{j \neq i}^{N} \min_{s} E(i_t, j_s)$$

where $E(i_r)$ is the side chain-backbone energy, and $E(i_r, j_s)$ is the side chain-side chain energy.

This algorithm results in the explicit enumeration of all the possible amino acid/rotamer pairs, leading to an exponentially increasing computational time for long sequences[43]. In order to encompass intrinsic limits of this search method, several expedients have been proposed to speed up the calculation. In particular, Generalized DEE compares clusters of rotamers instead of single rotamers, improving the convergence of the global search, which is also supported by a set of stochastic Monte Carlo trajectories[44].

Mean-field theory, rather than enumerating sequences, aims to calculate the site-specific average energy for each rotamer/amino acid combination, weighting the contribution of all the possible

combinations of the other residues by means of their probabilities. The general form of the average local energy $\varepsilon_i(\alpha_i)$ at position i is:

$$\varepsilon_i(\alpha_i, r(\alpha_i)) = \sum_{r(\alpha_j)} \sum_{j,\alpha_j} \omega_j(\alpha_j, r(\alpha_j))\, \gamma_{ij}(\alpha_i, r(\alpha_i); \alpha_j, r(\alpha_j)) + \varepsilon_i^0(\alpha_i, r(\alpha_i))$$

where $\gamma_{ij}(\alpha_i, r(\alpha_i); \alpha_j, r(\alpha_j))$ is the two-body interaction energy between side chains α_i and α_j in their rotamers $r(\alpha_i)$ and $r(\alpha_j)$; $\varepsilon_i^0(\alpha_i, r(\alpha_i))$ is the energy term for the single residue describing interaction with the backbone and its propensity for secondary structure. $\omega_j(\alpha_j, r(\alpha_j))$ is a Boltzmann term describing the probability for a certain amino acid to be at a given site[41]. All the field equations are solved and decreasing an effective temperature term identifies best energy solution. Mean-field algorithms demonstrated to be efficient in the design of hydrophobic cores, but less for solvent exposed positions.

To overcome these limitations the SCADS probabilistic approach generalizes the mean-field theory, orienting the search in the maximization of the total sequence-conformational entropy[43]. This approach has the advantage that several adimensional functional constraints can be added to guide the iterations toward convergence without an explicit weighing term for non-energetic contributions conversion. This algorithm, instead of giving the best sequence for a determined fold, results in a set of probabilities for a given amino acid to be at a certain sequence position. This represents a sensible information for the designer, since positions that are crucial for correct folding are easily identified[45].

Unfortunately, all the described algorithms cannot point out a single solution not only for the intrinsic limits of the algorithms themselves, but mostly because there may not be a single solution to

such a combinatorial problem. This is easily understood if we think about what nature does, preferring few folds and it can be visualized in the fitness surface mentioned earlier. Nature has found through evolution some robust tridimensional organizations for peptide chains, which can be encoded by sequences, sharing even very low amount of residues[46].

Further on, there are several well-established biases in super-secondary arrangements including packing of α-helices against β-sheets[47], shearing and twisting of β-sheets[48], β-turn, and α–α linking geometries[49].

Koga et al. adopted a computational approach to identify several structural rules for the stabilization of α–β folds[50], by means of *ab initio* folding calculations in the Rosetta suite environment. Zhang and Grigoryan, on the other side, adopted a statistical approach (Figure 6).

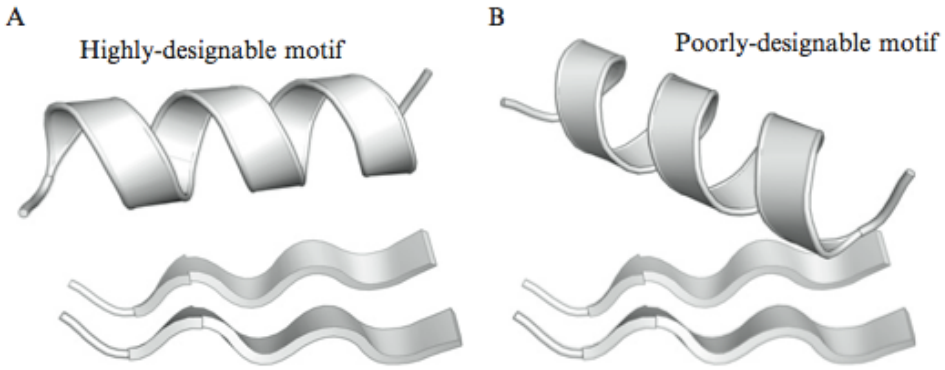


Figure 6. Example of designability of a β–β–α motif as assessed by MaDCaT. Adapted from reference [51].

Chapter 1: Introduction

Mining the ever-growing structural database, the MaDCaT software assesses the convenience of a certain tertiary motif, even suggesting sequence patterns that may encode it[51].

In the context of folding and inverse folding problems, this structure robustness in the sequence space has been referred to as designability[52]. The final aim of a protein designer is to *a priori* limit oneself to engineering only reasonably designable structure. It should be now clear that finding the relationship between structure and designability is a fundamental problem in protein design.

Designability concept in the context of four-helix bundles and coiled coils in general is a classical example of a structural motif with strong geometric preferences[53]. This leads to easy mathematical formulation, fully describing the global disposition of peptide chains, which are then crucial for the *de novo* design of proteins from scratch. Grigoryan and DeGrado demonstrated designability of coiled coils via a generalized mathematical model for backbone parameterization, producing deviation between ideal and real structures within 1 Å[53]. This model is the Crick parameterization of coiled-coils, which can be easily expressed as the relationship between the superhelical frequency ω_0 and the superhelical radius R_0 (Figure 7):

$$R_0 \cdot \omega_0 = d \sin(\alpha)$$

where ω_0 is in radians, α is the pitch angle, and d is the rise per residue.

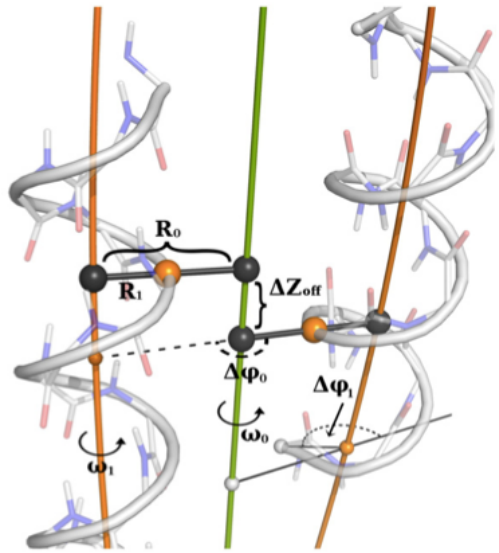


Figure 7. Visual represantation of parameters compounding the Crick fitting of coiled coils. Geometrical meanings of the superhelical radius (R_0), the helical radius (R_1), the superhelical frequency (ω_0), the helical frequency (ω_1), chain axial offset (ΔZ_{off}), chain superhelical phase offset ($\Delta\phi_0$), and starting helical phase ($\Delta\phi_1$) are shown. The green tube represents the interfacial axis. Orange curves depict local helical axes, which in a coiled coil form a superhelix. The gray tube represents the helical curve, which passes through Cα atoms. Orange balls show the inward-facing points on the helical curves (not necessarily corresponding to locations of atoms) defined as points with helical phase of π. The distance along the interfacial axis between an inward-facing point on one helix and its closest counterpart on the opposite helix is defined as ΔZ_{off}, with sign indicating the order of the two points, relative to N→C of the first helix. The depicted case is an anti-parallel coiled coil with a positive ΔZ_{off} of 2.4 Å. Adapted from reference [53].

The authors pointed out several key factors in the choice of the overall set of parameters, directing future works on the design of coiled coils and four-helix bundles, also providing a useful web-server for the automated fit and generation of coiled coils based on Crick parameterization, namely CCCP.

In the present work, parameterization of a four-helical bundle proteins housing a metal cofactor, was crucial.

1.2.1. Naturally occurring four-helix bundle metalloproteins

Lombardi et al in a seminal work made a retrostructural analysis of metalloproteins[54]. In particular, they focused on diiron oxo binding sites, housed in a four-helix bundle. They dissected the main factors, which contribute to define the structure and geometry of the metal-binding site in this class of proteins. They analyzed ferritin[55], bacterioferritin[56], rubrerythrin[57], ribonucleotide reductase R2 subunit (R2)[58], Δ9 ACP desaturase[59], and the catalytic subunit of methane monooxygenase[60], a member of the Bacterial Multicomponent Monooxygenases (BMMs) superfamily[61]. Even though these proteins show less than 5% sequence identity, their active sites are housed within a very simple C_2 symmetric four-helix bundle. It was pointed out that a simple model for the four-helix bundle in terms of a D_2 symmetric tetrameric coiled coil could satisfy the structural features of the natural proteins, with a slight RMSD deviation around 1 Å (Figure 8).

Protein name	rms deviation (Å)	Atoms superimposed	Symmetry	α, degrees	β, degrees	γ, degrees	Xtrans, Å	Ytrans, Å	Ztrans, Å
Bacterioferritin[c]	1.03	48	D2	8.5	13.6	−4.1	4.5	5.6	7.5
Rubrerythrin[d]	0.89	48	D2	10.3	11.2	−3.7	4.2	6.0	7.3
H Ferritin[e]	1.37	48	D2	6.7	8.5	−4.1	4.3	6.3	8.0
Ribonucleotide reductase (R2)[f]	1.11	48	D2	3.9	18.3	−1.4	4.6	5.7	6.9
Methane monooxygenase[g]	0.96	36	D2	0.1	19.9	2.2	4.1	5.4	6.5
Δ9 ACP desaturase[h]	1.29	36	D2	−11.9	14.8	−2.2	4.5	5.7	9.1
Myohemerythrin[i]	1.23	74	D2	4.8	10.0	4.5	5.3	5.2	7.3
Hemerythrin[j]	1.45	74	D2	0.2	9.6	3.7	5.2	5.2	7.2
Hemocyanin[k]	2.09	42	D2	5.5	14.4	−17.1	6.9	6.1	11.2

Figure 8. Retrostructural analysis of diiron four-helix bundle proteins as taken from reference[54]. Enforced D2 symmetry satisfies the arrangement of the backbone in the four-helix bundle. Parameterization of the bundle is based on a method described in reference[22].

These proteins share the so-called ExxH motif, where a glutamate residue and a histidine residue, involved in the metal binding, occupy

the *a* and *d* positions of the coil, respectively. This binding motif is the key primary motif at the basis of this project and will be recursively found in this work.

Nature selected this four-helix bundle scaffold and this ExxH motif because of their ability to fine-tune the chemistry of the metals by means of i) the global orientation of the bundle, ii) the choice of the residues in the first coordination sphere, iii) the choice of the interacting residues of the second coordination sphere (proton shuttling, hydrophobic environment, pK_a of the ligands, redox properties of the metal) iv) the ability to bind exogenous cofactors and substrates.

All of these features can be found in the carboxylate-bridged diiron class of proteins. In the majority of these enzymes, the dimetal sites are bridged by a combination of oxo, hydroxo, or carboxylate donors; two histidine and four carboxylate ligands represent a mostly conserved protein-derived ligand set. The only exception is hemerythrin (Hr), which contains in its active site a high number of His ligands[62]. The functions of diiron carboxylate proteins range from dioxygen transport and activation, to phosphoryl transfer, to iron storage[25]. Several proteins pertaining to this class have been characterized and are still object of strong interest in the science community. Among them, relevant are: soluble methane monooxygenase (MMO) and the growing subclass of BMMs (including toluene monooxygenases, phenol hydroxylase, alkene monooxygenase), which will be described in more details in the next paragraph; the ribonucleotide reductase R2

subunit (RNR-R2), which generates a tyrosyl radical essential for the reduction of ribonucleotides to deoxyribonucleotides in DNA biosynthesis[63]; the stearoyl-acyl carrier protein (ACP) Δ9-desaturase, which introduces a double bond into saturated fatty acids[64]; ferritins and bacterioferritins, which use iron as a substrate for ferroxidation and iron storage[65]; Hr and myohemerythrin, which reversibly bind and transport oxygen[62]. More recently, the diiron carboxylate protein family broadened to include four membrane-associated enzymes, identified on the basis of six conserved amino acids, which constitute the iron binding motif. These proteins are alternative oxidase (AOX), plastid terminal oxidase (PTOX), 5-demethoxyquinone hydroxylase (DMQ hydroxylase), and Mg-protoporphyrin IX monomethilester hydroxylase (MME hydroxylase)[66].

The functional differences between all these proteins can be ascribed to a different mechanism in their interaction with dioxygen. In fact, most of these enzymes catalyze redox reaction by the use of dioxygen, binding it initially in the diferrous (bis-Fe^{2+}) state. The next step is the formation of diferric peroxide intermediate whose fate depends on the specific environment of the diiron site[67].

Other classes of proteins use a four-helix bundle as the protein moiety binding the metal cofactor. For example, the heme cofactor, of paramount importance in several biological processes, can be found in a four-helix bundle context, where it is bound in the hydrophobic cleft. In particular, cytochromes *c'* [68], which are found in photosynthetic bacteria, still serve an unknown function. Unlike Class I c-type

cytochromes, these cytochromes are high-spin: the iron is five-coordinate, with only one His as axial ligand. The vacant axial coordination site is buried in the four-helix bundle core and may serve as a redox potential regulatory sixth ligand position[69]. Another heme protein characterized by a four-helix bundle is the cytochrome b_{562}[70]. The *b*-type heme is oriented parallel to the bundle axis, with the two axial ligands being a histidine and a methionine residue. The heme environment is asymmetrical as the histidine bound face of the heme is exposed to the solvent, while the other is deeply buried in the four-helix bundle core. The simple fold of this protein made it the perfect candidate for several electrochemical and protein folding studies as demonstrated by the extraordinary work of Gray group[71]. Membrane protein cytochrome bc_1, which is involved in the Q cycle, houses two tightly bound hexacoordinated heme iron in a membrane inserted four-helix bundle[72]. Second coordination shells comprise two threonine residues hydrogen bonding the histidine residues in the fifth and sixth positions of the first coordination shell. This bundle is highly symmetric and has been proved useful for further protein design, as shown later.

1.2.1.a. The Bacterial Multicomponent Monooxygenases

Monooxygenases are key enzymes in the aerobic bacterial degradation of aromatic hydrocarbons. They catalyze the hydroxylation of organic compounds (methane, aromatic rings, unsaturated hydrocarbons) at different positions with high

regiospecificity[73]. This class of proteins is divided into six distinct groups, which differ for substrate specificity and mechanism of substrate conversion[73]. Among them, the most studied are undoubtedly methane monooxygenases (MMO) with their soluble[74] (sMMO) and unsoluble[75] (pMMO) counterparts, toluene monooxygenases (TMO) and their subfamilies[76] (T4MO, T3MO, ToMO based on their hydroxylation position) and phenol hydroxylases[77] (PH). Most BMMs, including the sMMO and PH, need three proteins to accomplish their function: a 200-225 kDa hydroxylase organized as an αβγ dimer housing the carboxylate-bridged diiron center, a 10-16 kDa cofactorless regulatory protein involved in the enhancement of the catalyitic activity synergistically guiding the events of the electron transfer and catalysis, and a 38-40 kDa [2Fe-2S]- and FAD-containing reductase that shuttle electrons in order to restore the metal catalyst (Figure 9). TMOs needs a fourth component for function, a 10-12 kDa Rieske protein mediating the electron transfer from the reductase to the hydroxylase[74].

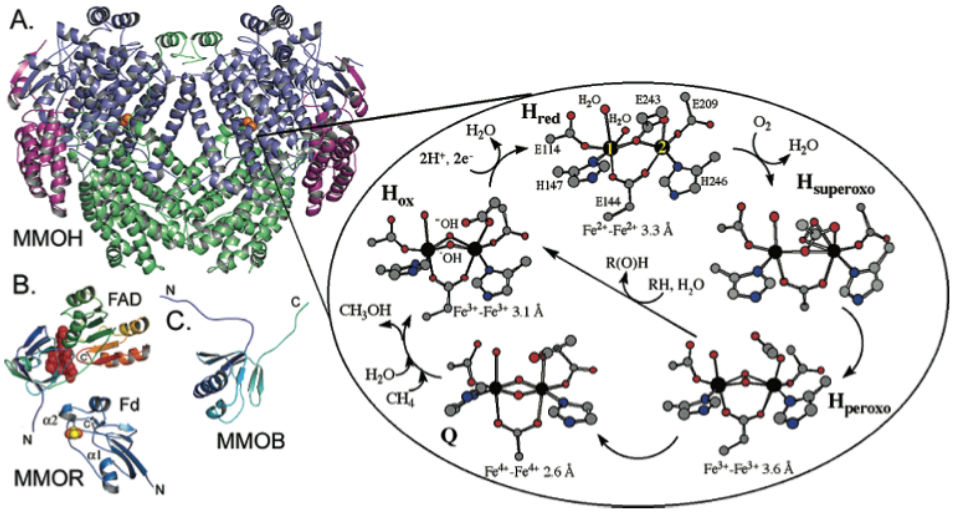


Figure 9. Structures of sMMO components and proposed reaction cycle. A) MMOH hydroxylase component; B) MMOR the reductase component; C) MMOB the regulatory component. Adapted from reference [61].

Crystallographic resolution of the hydroxylase component structure of several members of this family evidenced a similar global topology[78]. The 2-fold symmetric αβγ dimer forms a solvent exposed cleft at the interface between the α and β subunits, which has recently been shown to host the regulatory protein[79]. Helices B, C, E and F form the cofactor binding four-helix bundle in the α subunit. Helices E and F are solvent exposed, and are involved in the binding of the regulatory component, which dramatically stabilize the conformation of residues crucial in the proton shuttling to the active site[80].

Among the BMMs, the most studied is sMMO. The catalytic chemistry of this system is extraordinary, since it can catalyze the cleavage of one of the most stable bond in nature, the C-H bond in

methane, promoting the formation of methanol according to the reaction scheme:

$$CH_4 + O_2 + NADH + H^+ \rightarrow CH_3OH + NAD^+ + H_2O$$

In all these proteins, the ligating modes of the active site carboxylate ligands have been shown to be flexible, giving rise to a carboxylate shift phenomenon, occurring upon changes in the oxidation state of the diiron site.

The MMOH dimetal site, rich in carboxylates, tends to consume oxygen rather than binding it reversibly. During the catalytic cycle, a variety of intermediates has been proposed and numerous spectroscopic techniques and theoretical calculations have been used to identify these compounds[81]. MMOH binds O_2 in the diferrous state, leading to a diferric intermediate with a symmetrical bridging peroxo group[81]. Then, the site evolves through several reactions, giving rise to a reactive high-valent iron-oxo intermediate diferryl (bis-Fe^{4+}), named compound Q[82] (Figure 10).

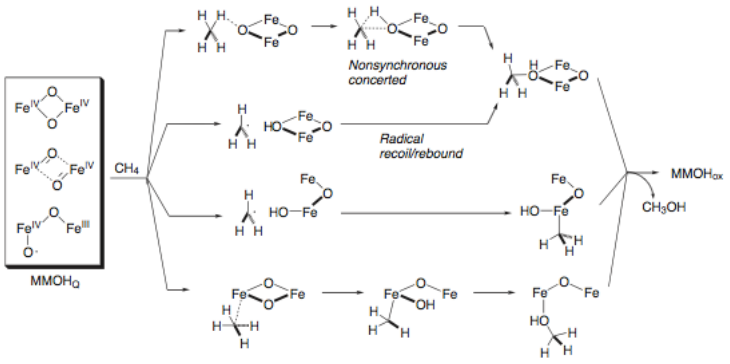


Figure 10. MMOH compound Q geometry s been proposed in several isomeric forms by DFT calculations. Calculations for the preferred mechanism lead to four plausible pathways. Adapted from reference[82].

In all the structures of MMOH oxidized diiron(III) form, both iron ions have a six-coordinate distorted octahedral environment (Figure 9 and Figure 11), and reside in the center of a four-helix bundle, which provides two ExxH motifs[25]. Fe1 is coordinated to the δ_1-nitrogen atom of a histidine residue, His147, a monodentate carboxylate, Glu114, and a terminal water molecule. Fe2 is ligated by His246, two monodentate carboxylates, Glu209, and Glu243. The terminal water molecule coordinated to Fe1 forms a hydrogen bond with the uncoordinated oxygen atoms of both Glu114 and Glu243.

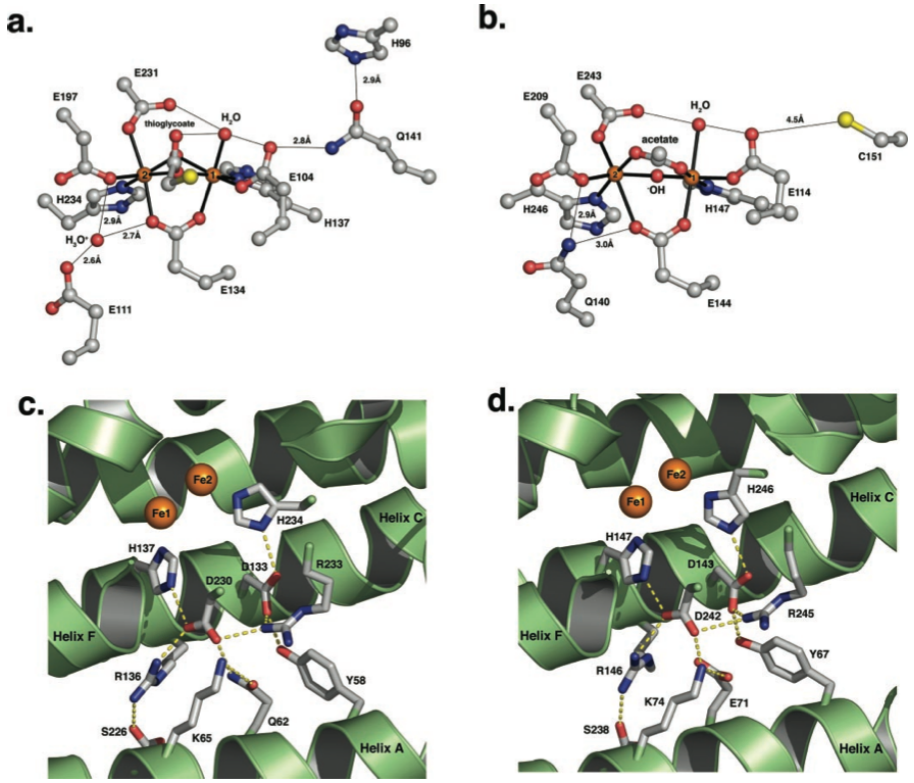


Figure 11. Hydrogen bond network in the first (a, b) and second (c, d) coordination shell. Note the almost identical interaction patterns for MMOH (b, d) and ToMOH (a, c). Adapted from reference [78].

The two iron ions are bridged by Glu144, a hydroxide ion and a third ligand that is variable in the different crystal structures. Depending on the crystal form, the species and the temperature of data collection, bridging water, hydroxide, acetate, and formate have been reported as the third ligand[74,83]. Changes in the structure occur upon reduction of the binuclear iron cluster to the diferrous [Fe^{2+}-Fe^{2+}] form (Figure 9). One specific ligand, Glu243, modifies its coordination mode, giving rise to the so-called "carboxylate shift." Glu243 displaces the bridging hydroxide ion and becomes a bidentate chelating ligand to Fe2, while it forms a monodentate interaction with Fe1. As a consequence, both iron ions become five-coordinated. The ligand flexibility allows ligands to coordinate the Fe centers differently, depending on the enzymatic needs: as bidentate ligands (bridging or chelating), when saturation of the first coordination sphere of Fe ions is required, and as monodentate (terminal) ligands, when one (or more) vacant coordination site(s) in the Fe center are needed (e.g., upon dioxygen or substrate coordination) to enable a certain reaction step to take place. The consequence of these structural changes is that both five- coordinated iron(II) ions react with O_2.

The diiron center is located on the surface of a hydrophobic cavity, which mainly contains hydrophobic side chains, with the exception of Cys151 and Thr213. The occurrence of Cys151 at this location suggests a possible role for the sulfhydryl moiety as a one-electron source and/or as a proton donor. The threonine may also participate in the oxidation reaction by functioning as a proton source. Such a role

for Thr213 would be mechanistically analogous to that of Thr302 in the related heme oxygenase, cytochrome P450[84]. Furthermore, a hydrogen bonding network between residues belonging to the second coordination sphere, is observed (Figure 11). The network extends from the two iron-coordinating histidines, His147 and His246, to the surface of the active site. The involved residues are Asp143 and Asp242, each of which is hydrogen bonded to the NεH of the histidine ligands[61], and to surface residues, Tyr67 and Lys74.

Given the simple organization of the backbone as shown previously, the asymmetry of the interaction generating the coordination imbalance between the two iron atoms may delineate a scenario in which structural requirements for diiron binding are fulfilled by the symmetry of the four-helix bundle. However, the catalytic activity is reached by the plasticity of the coordination site, which is further stabilized by the asymmetrical second coordination sphere.

1.2.2. *De novo* designed four helix-bundle metalloproteins

First metalloprotein designs started as soon as the principles and methods for designing well-defined protein were formulated[9]. Metalloprotein design may be considered more ambitious, since one has to fully control the structure of the protein scaffold as well as the metal site in the context of an artificial protein. As pointed out by the analysis of the metal binding site in BMMs, a plethora of interactions are responsible for finely tuning the metal site activity in

metalloproteins[25]. As a consequence, to construct an artificial metalloprotein several issues must be simultaneously satisfied: the design of a correct protein structure and the metal ion coordination requirements. As a result, metalloprotein design is a complex task and different design approaches have been adopted. They span from the binary patterning of hydrophobic and hydrophilic residues, to the totally automated computational approach[9]. Due to the complexity of the problem, cycles of design, synthesis, characterization, and redesign are often needed to get a protein with all the desired features[85].

Several structural motifs have been adopted for metalloprotein design, such as β-sheets, α/β-motif, but α-helices and in particular helical bundles almost monopolized the attention of researchers in these more than 20 years of metalloprotein design [9,36–38]. A complete review of all the astonishing results in this field may be found in comprehensive books[86]; here the focus is on four-helix bundle based metalloprotein design, with a particular sight on dinuclear metalloproteins.

Pioneering work by Handel and DeGrado opened the way to many other protein designers in the exploitation of the four-helix bundle for metal binding purpose. Their α4-derived designed protein formed a Zn^{2+} three-coordinate site, consisting of three histidine side chains as confirmed by NMR spectroscopy[87]. The same fold was also used as a template for engineering a tetrahedral Cys_2His_2 Zn^{2+}-binding site, similar to the zinc finger proteins[88]. Ala-scanning of the coordinative residues demonstrated to influence the metal binding affinity[89].

Tanaka's group accomplished in the *de novo* design of type 1 blue copper protein, called AM2C[90]. They designed a two His, one Cys Cu(II) binding site on the crystal structure of a tetrameric mutant of the GCN4 Leu zipper. To express the asymmetric four-helix bundle, three Gly-rich loops were added. The AM2C folding, metal-binding properties, and coordination geometry were ascertained by several spectroscopies, and the metal binding site geometry was confirmed by EXAFS analysis. The protein showed a good folding even in the absence of the metal cofactor. The Cu(II) ion is coordinated with a distorted trigonal geometry by the endogenous ligands, while a fourth coordinative site is filled by chloride ion. Despite the structural differences, AM2C showed electrochemical properties similar to azurin.

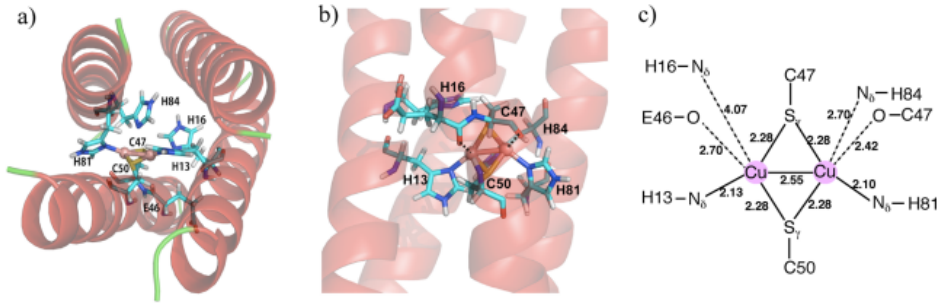


Figure 12. bi-AM2C di-Cu(II) cofactor. a) top and b) side view of the full protein model. c) shows the proposed metal binding geometry consistent with the EXAFS data. Adapted from reference [91].

Recently, a second redesign of this protein led to a purple copper site similar to the Cu_A site of the cytochrome c oxidase. This protein, called bi-AM2C, was obtained by simply adding three more coordinating residues in the core of the four-helix bundle, in a CxxC,

and two HxxH motifs. Exhaustive spectroscopic analysis confirmed the presence of a di-Copper cofactor, with the two Cys residues in bridging positions; two tightly bound His and four labile N/O donors from histidines and backbone, respectively (Figure 12). The purple color of this designed metalloprotein faded out in few hours, undergoing to an oxidative quenching of the cysteines residues.

Ogawa and coworkers reported the assembly of a four-stranded coiled coil upon binding of Ag(I) and Cu(I), in the series of peptides called C16C19-GGY. These soft metal ions drive the holigomerization by the formation of a tetranuclear metal ion thiolate cluster[92]. The same behavior was not observed with harder metals like Fe(II) and Zn(II). The multinuclear Cu(I) site is responsible for an intense room temperature long-lived luminescence, centered at 600 nm, as found in several polynuclear Cu(I) cluster natural proteins. Luminescence may be quenched upon addition of oxidants, reductants and urea, confirming the importance of the copper oxidation state and its burial within the protein, which prevents solvent quenching. EXAFS spectroscopy suggests that only four out of eight Cys are involved in the metal binding and that N/O ligands are involved in the stabilization of the terminal coordinative sites.

Nanda and coworkers computationally designed a four-helix bundle protein, CCIS1 (Coled Coil Iron Sulfur protein 1), capable of incapsulating in its hydrophobic core a Fe_4S_4 cluster[93]. The significance of this design is in the α-helical scaffold adopted in the design, atypical for the natural counterparts, which are rich in β-structures. The Fe_4S_4

cluster is inserted in the core of the protein, where four C_dxxxC_a motifs create a D_2 symmetrical binding environment. The automated design approach retraces the method adopted in the work on diiron metalloproteins, discussed later in the text. Results on CCIS1 were obtained by several spectroscopic analyses and are particularly promising for future design of multi-cluster wire for bioelectronics purposes.

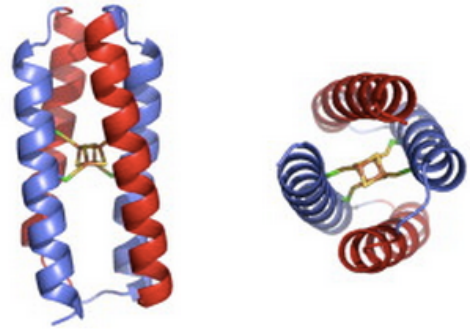


Figure 13. Side (left) and top (right) views of the CCIS1 Fe4S4 designed model. Adapted from reference[93].

Another prolific field of *de novo* design is represented by the heme-binding metalloproteins. DeGrado, Dutton, and coworkers pioneered this field in a series of works in which they based the design on a previously designed α_2B sequence[94,95], a helix-loop-helix peptide that dimerizes to form a four-helix bundle. The design process led to two sequences, named $VAVH_{25}$(S-S) and Retro(S-S), both possessing a N-terminal disulfide bridge to ensure the dimeric oligomerization. Retro(S-S), the reversed sequence version of $VAVH_{25}$(S-S), binds the ferric heme cofactor with higher affinity with respect to $VAVH_{25}$(S-S), and both show the correct 1:1 stoichiometry. Even though several

spectroscopic data indicated the correctness of the design, the negative reduction potential indicated that heme is partially exposed to the solvent.

Dutton and coworkers obtained outstanding results, with their *maquettes*, α-helical peptides with the heptad repeat pattern typical of left-handed coiled coils with the ability to bind arrays of cofactors and reproducing native-like functions. Work made on Dutton's *maquettes* is a successful example of the rational redesign approach. Starting from the first model H10H24, they reached the aim of obtaining the expected binding ratio of four hemes in an $(\alpha\text{-SS-}\alpha)_2$ motif[94]. Successive redesign cycles[96] allowed to delineate some key rules for the binding and tuning of the heme cofactor: i) the designed scaffold should provide an highly hydrophobic environment for tight binding of the heme cofactor, ii) to attain pH redox-dependent behaviors, the scaffold must retain the fold upon side-chain protonation changes, iii) the burial of protonatable groups into the hydrophobic protein interior could modulate the pK values of acidic side chains, thus allowing redox-coupled proton exchange, iv) modulation of heme redox potential can be achieved by charge compensation: negatively charged residues favor ferric heme, thus determining negative redox potential values, v) since charge compensation effects are distributed between many groups, the design of hemeprotein with pH-independent redox behaviors can be accomplished by removing all charged groups in the proximity of the heme. Several cycles of further redesign in order to gain a less flexible structure[85] led to the design of the HP1, which is

endowed with all the achievements reached with its precursors. HP1 binds two hemes per four-helix bundle with K_D values < 20 nM. But only with HP7, the bis-His ligated hemes showed oxygen transport properties[97]. This maquette demonstrated to undergo to allosteric and cooperative regulation of the oxygen binding as in the natural proteins. Further on, they showed that the exclusion of water from the protein interior, confirmed by hydrogen/deuterium exchange by NMR, was crucial to preclude heme oxidation. The spectral properties of this analogue are very similar to natural neuroglobins, and CO binding affinity was lower than O_2, suggesting that distal histidine may have a role in stabilizing the dioxygen binding.

DeGrado and coworkers adopted a computational *de novo* design of heme-binding four-helix bundle[98]. Adopting the backbone parameterization discussed earlier, they afforded the design of a cytochrome bc_1 analogue with a completely automated design approach, which is based on the D_2 four-helix bundle. A 25-residue long peptide was designed with a His at *d* position and a Gly at *a* position on the other side of the bundle, thus binding two heme cofactors. A Thr residue is involved in a second-shell interaction with the binding histidine as found in cytochrome bc_1. SCADS algorithm was adopted to pack all the left residues. Design then turned to the design of a peptide, which could be able to bind an abiological heme binding as the diphenylporphyrin[99,100]. PA_{TET} and its single chain counterpart PA_{SC} both bind the cofactor specifically and with the correct stoichiometry. Later they designed a heterotetramer able to

specifically bind two Zn(II)-DPP instead of its ferric form[101], resulting in fluorescent adduct. This accomplishment was made possible by deleting one axial His residue with respect to PA_{TET}.

All these examples demonstrate how a *de novo* designed protein is not a sterile mind exercise, but also a valuable test of our folding and metal binding knowledge, as well as an instrument to always learn more lessons from nature, through these bio-inspired models, reproducing native-like functions.

1.2.2.a. The DF family

Starting from the parameterization of the natural diiron metalloproteins, mentioned before[54], the Artificial Metallo Enzyme Group at the University of Napoli "Federico II", in collaboration with the DeGradoLab at Penn University, Pennsylvania now moved to University of California, San Francisco (UCSF), followed a *de novo* design approach for the development of the DF (Due Ferro) family of artificial proteins.

Design started taking advantage of the pseudo C_2 symmetry of this class of proteins. This symmetry has been shown to be particularly advantageous simplifying the design and their structural characterization, and reducing the size of the synthesized peptide. The first member of this family, DF1, is made up of two 48-residue helix-loop-helix (α_2) motifs, autonomously self-assembling into an antiparallel four-helix bundle. To provide the ligands of the first coordination shell, two helices bear the ExxH motif, as found in

natural proteins, where a bridging Glu residue is in *a* position and a terminal His is in *d* position of the coiled coil describing the four-helix bundle. A second Glu residue is placed on the other two helices, providing a fourth ligand for each metal ion. Other keystone residues of the second coordination shell are intended to stabilize the polar metal binding site in the middle of the protein as in the natural counterparts. Thus a Lys/Asp/His H-bond network was designed similar to the Arg/Asp/His pattern found in BMMs, while a Tyr residue donates a proton to the non-bridging carboxylate ligand (Figure 14). Each metal is five-coordinated in order to host an exogenous ligand.

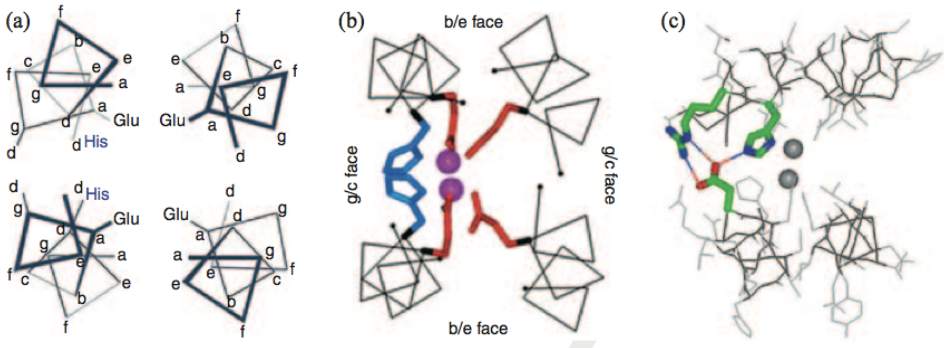


Figure 14. Tetrameric coiled coil heptad repeat adopted in DF1 (a); metal binding site Glu and His residues in red and blue sticks respectively (b); His/Asp/Arg H-bond network as found in natural proteins (c). Adapted from reference [25].

All of the residues in the core and those in contact with the solvent were placed in order to satisfy perfect packing and the correct antiparallel topology. Finally, an idealized γ–α_L–β interhelical loop was included to stabilize the α_2 motif.

Chapter 1: Introduction

DF1 was able to bind Zn(II), Co(II), Mn(II), or Fe(II) in the micromolar range. DF1 demonstrated to be very stable, and the design correctness was confirmed by the crystal structures in complex with Zn(II), Co(II) and Mn(II)[54,102], as well as the solution NMR structure of the *apo*-form confirming that the metals bind the protein in a pre-organized environment as in the natural proteins[103]. DF1 is the first protein, entirely designed from scratch, for which both crystallographic and NMR structures have been obtained.

Although DF1 behaves as a native-like protein, as it was designed for maximal stability, it doesn't show any functionality, and its solubility is limited.

Table 1-1. Peptide sequence of DF family. Liganding residues are in bold, helical regions are underlined, active site access residues are in italic.

			Sequence						
	g	abcdefg	abcdefg	abcdefg	abcdefg	abcdefg	abcdefg	abcdefg	abcdef
DF1		DY	LRELLK*L*	**E**LQ*L*IKQ	YREALEYV--KL-PV	LAKILED	**E**EK**H**IEW	LETILG	
DF3		DY	LRELLK*G*	**E**LQ*G*IKQ	YREALEYT--HN-PV	LAKILED	**E**EK**H**IEW	LETILG	
DF2	M	DY	LRELYK*L*	**E**QQ*A*MKL	YREASERV--GD-PV	LAKILED	**E**EK**H**IEW	LETING	
DF2t	M	DY	LRELYK*L*	**E**QQ*A*MKL	YREASEK-ARNPEKKSV	LQKILED	**E**EK**H**IEW	LETING	
$DF1_{sc}$(1-28)		DE	LRELLK*A*	**E**QQ*A*IKI	YKEVLKKA--KE-GD				
…(29-59)		EQELARL	IQEIVK*A*	**E**KQ*A*VKV	YKEAAE----KA-RN				
…(60-88)		PEKRQV	IDKILED	**E**EK**H**IEW	LKAASK-----Q-GN				
…(89-114)		AEQFASL	VQQILQD	**E**QR**H**VEE	IEKKN				
3H-DF_{sc}(1-28)		DE	LRELLK*G*	**E**QQ*G*IKI	LKEVLKKA--KE-GD				
…(29-59)		EQELARL	NQEIVK*G*	**E**KQ*G*VKV	YKEAAE----KA-RN				
…(60-88)		PEKRQV	IDKILED	**E**EK**H**IEW	HKAASK-----Q-GN				
…(89-114)		AEQFASL	VQQ**H**LQD	**E**QR**H**VEE	IEKKN				
DFtet: A	K	LKELKSK	LKELLK*L*	**E**LQ*A*IKQ	YKELKAE	LKEL			
Aa	E	LKELKSE	LKELLK*L*	**E**LQ*A*IKQ	FKELKAE	LKEL			
Ab	K	LKKLKSR	LKKLLK*L*	**E**LQ*A*IHQ	YKKLKAR	LKKL			
B	E	LEELESE	LEKILED	**E**ER**H**IEW	LEKLEAK	LKEL			

In the first round of redesign, DF1 and DF2 subsets of analogues were mainly devoted to overcome the limits of DF1 in active-site accessibility and solubility (Table 1-1). Substitution of Leu residues with smaller Ala and Gly, afforded a cavity large enough to allow the

access of small molecules. L13A-DF1 and L13G-DF1 are able to bind oxygen displaying ferroxidase activity, azide, phenol and acetate, and their X-ray structures were solved, displaying a great flexibility in the metal binding geometries able to accommodate up to two water molecules per dinuclear site[102]. DF2 variants improved the total solubility[104] and explored the role of the loop in the folding and metal binding properties[12]. To optimize the scaffold for catalytic activity an approach more similar to a "combinatorial" screening was needed.

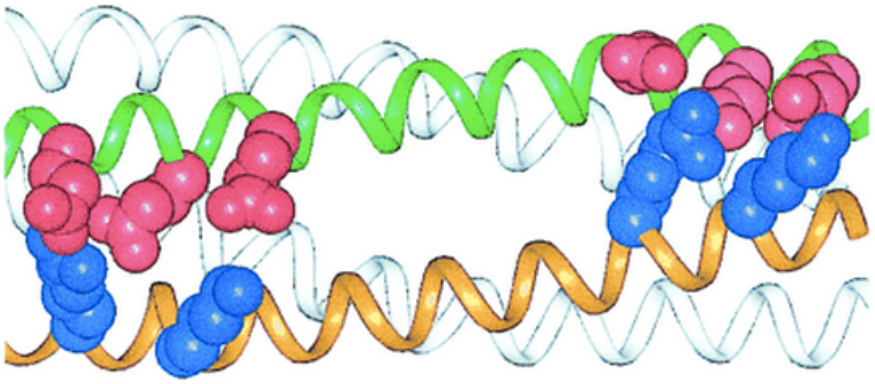


Figure 15. Polar sidechains to drive the correct antiparallel pairing of the coiled coil in DF_{tet} subset. Adapted from reference [105].

The DF_{tet} subset (Table 1-1) proposed to fill the gap between "rational design" and "combinatorial" approaches by means of a four-helix bundle heterotetrameric system[33]. First an A_2B_2 system and then a more generalized $A_aA_bB_2$ system[105] would give raise to a thousand of different compounds starting from only 10 variants of each component. In order to stabilize the antiparallel coiled-coil the 24-long helices in DF1 were extended to 33 residues, offering an increased hydrophobic interface to drive the folding (Figure 15). Next a course-

grained energy function accounting only for electrostatic interaction drive the design of the solvent exposed region in a way that could stabilize only the correct topology and destabilize undesired parallel tetramerization. This design was the first reported in which an explicit "negative design" was accomplished. The resulting protein showed the correct oligomerization by CD spectroscopy, size exclusion chromatography and analytical ultracentrifugation, as well as binding of Co(II), Zn(II), and Fe(II). As in the DF1 subset a series of asymmetrical variants were synthesized to explore the ferroxidase activity as function of the accessibility of the metal binding site[106]. G_4-DF_{tet}, where four leucines in the proximity of the metal site were replaced by four glycines, exhibited the most rapid rate of iron oxidation, with no detectable intermediates. This analogue was able to enhance ≈ 1000-fold the reaction rate of the two-electrons oxidation of 4-aminophenol to immino quinone. DF_{tet} subset clearly provides precious insights in the active site requirements for function. The side chain hindrance around the metal greatly influences the catalytic efficiency of the protein, and the combinatorial approach demonstrated to be very promising for identifying candidates for different applications.

Lessons learned from the DF_{tet} subset were applied either into a single-chain protein or into the well-characterized helix-loop-helix scaffold.

A 114-residue protein, DF_{SC}, was designed adopting the helices of DF_{tet} as a template four the bundle[45]. The keystone residues of the first

and second coordination sphere as well as the loop regions were kept fixed, while SCADS repacking algorithm guided the choice of the other residues. The resulting sequence was highly soluble and expressed in *E.coli* in high yields. DF_{SC} binds Zn(II), Co(II), Fe(II), Mn(II) in the intended stoichiometry with micromolar affinities. A QM-MM refined NMR structure was solved and proved to be in good agreement with the starting model. This protein presented four Ala residues at the entrance of the metal binding site, but it was not able to convert any substrate, probably because of the low flexibility of the scaffold.

Later, with the aim to circumvent these problems and to simplify the structural analysis, a new variant was synthesized adopting the same helix-loop-helix motif of DF1. Four Gly residues were placed for active site accessibility as found crucial in DF_{tet} (Figure 15).

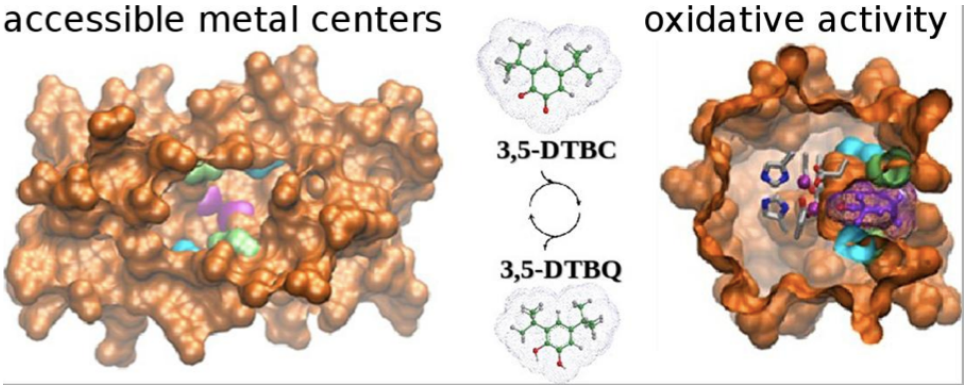


Figure 16. DF3 open active site makes it able to reach oxidase like activity. Adapted from reference [107].

Since only one Gly destabilizes DF1 by 10.8 kcal/mol[108], the interhelical loop was carefully redesigned in order to overcome the

instability induced by the introduction of destabilizing residues. The original Val[24]-Lys[25]-Leu[26] of DF1 was changed to Thr[24]-His[25]-Asn[26]. The newly designed DF3 exhibited enhanced solubility (up to 3 mM in water) and active site accessibility, while retaining the intended global fold in the holo form as confirmed by the NMR solution structural characterization. DF3 was the first *de novo* designed enzyme, as it is able to catalyze the two-electron oxidation of the substrate 3,5-ditert-butyl-cathecol to quinone with a saturation kinetics typical of natural enzymes.

Recently, a redesign process of the DF_{SC}, led to two new variants: G_4-DF_{SC} and 3His-DF_{SC}[109] (Figure 17). Both variants exhibited catalytic activity, but an automated design process induced a different reactivity, as the first is able to oxidize 4-aminophenol to quinoneimine as in DF_{tet} while the second promoted the N-hydroxilation of p-anisidine.

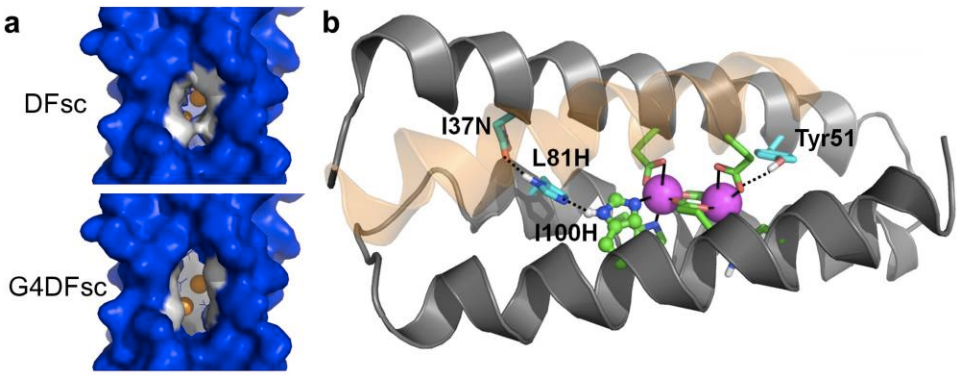


Figure 17. G4DFSC presents four Gly residues, allowing substrate access (a); 3His-DFSC H-bond network designed to stabilize the additional His binding residue in position 100 (b). Adapted from reference [109].

The introduction of an additional His in the first coordination shell was inspired by the natural protein AurF, which exhibit a similar reactivity. Interestingly the automated force-field oriented design gave the same hydrogen bond pattern as found in the natural protein to stabilize the insertion of the polar residue in the hydrophobic core of the protein. However both the proteins were not able to catalyze these reactions as they are able to make only a single turnover, moreover both of them and in particular 3His-DF_{SC} do not show a regular folding in their apo-forms.

The DF family experience undoubtedly demonstrated the uniqueness of the four-helix bundle scaffold in affording the difficult task to host a metal cofactor and the tunable structural flexibility needed to exert difficult chemical transformations.

1.3. Expanding the scope of possible function in DF proteins

In this work we aim to open new perspectives in the designable space of DF family. The goal was to improve our knowledge of the basic principles regulating folding and metal reactivity, while simultaneously design basic biochemical components and nanodevices for analyte sensing, chemical synthesis, electron exchange/storage towards solar energy conversion[110]. We have adopted a dual approach, leading to the *de novo* design of two new subsets of the DF family. Symmetric properties of the four-helix bundle scaffold have been tested in order to understand the role that symmetry exerts on the chemistry

of the metal cofactor. To this end, we eliminated on the one side the C_2 symmetric constraint of the DF1 scaffold with the DF-Click subset, while, on the other side, we reinforced this geometrical constraint designing a new class of metalloproteins, the QFs (Figure 18).

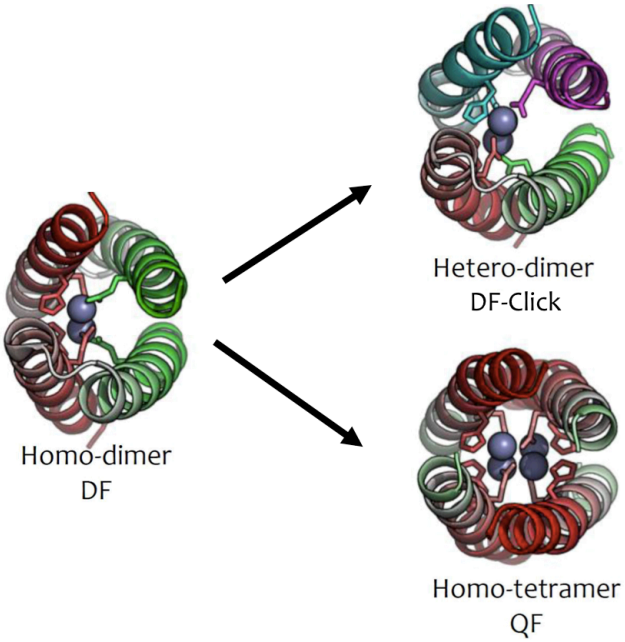


Figure 18. Top view of the DF1 dimer on the left, and the two strategies adopted in this work on the right.

As showed above, the asymmetrical members of the DF are successful examples of how an asymmetrical environment around the dinuclear cofactor induces a shift in its catalytic properties[106,109]. More in general, it is well established that fine-tuning of the second shell in an asymmetrical context is crucial for metal specificity[69], reactivity and redox properties of the cofactor[111], as well as specificity for the substrate[83] in natural proteins, and shown by BMMs.

On the other site, highly symmetric systems proved to be functional in many biotechnological applications, Ueno and coworkers made brilliant work in the redesign of icosahedral bacterioferritin as molecular cage for several purposes[112], while Woolfson and coworkers adopted long-range coiled coil association to generate new materials as hydrogels[113] and nanoparticles[114], Yeates and coworkers exploited the metal directed self-association of redesigned cytochrome b_{562} to induce pH dependent crystallization[115], DeGrado and coworkers made screw-symmetric bio-nanocomposites from the association of highly organized α-helices on a carbon nanotube[116], and DeGrado, Saven and coworkers made the first *de novo* design of a protein crystal creating the foundations on which this work is placed[117]. Here, we want to exploit the opportunity given by symmetry of generating superstructures in a controlled fashion.

1.3.1. DF-Click: design of asymmetric metal binding sites

Although the DF_{tet} and DF_{SC} subsets undoubtedly reached outstanding results, both strategies present some intrinsic limits. Indeed, as the ease of synthesis and combination of new analogues is the strength of the former strategy, it is the drawback of the latter. Vice versa the structural robustness of the latter is the main limit of the former strategy.

For this reason we developed a new strategy for the synthesis of heterodimers. We adopted the DF1 α_2 motif as it is more reliable in terms of ease of analysis and structure stability, and to be more

amenable to a rational/combinatorial approach as it can be chemically synthesized.

An innovative synthetic approach was developed in this work by means of the so-called "Click Chemistry", which have ensured ensure the formation of the desired heterodimer. Our design strategy led to helix-loop-helix heterodimer obtained by side-chain chemical ligation, the DF-Click series.

Here is reported the design, synthesis and spectroscopic characterization of the first analogue of this series named DF-Click1, which was inspired by the hydroxylase component of the BMMs. We proof the correctness of our design process and its applicability to the creation of a complete library of ligands, in order to gain new functions. Thanks to the synthetic protocol developed for covalent binding, the suitability of this approach in discriminating among the possible undesired dimer combinations is further discussed.

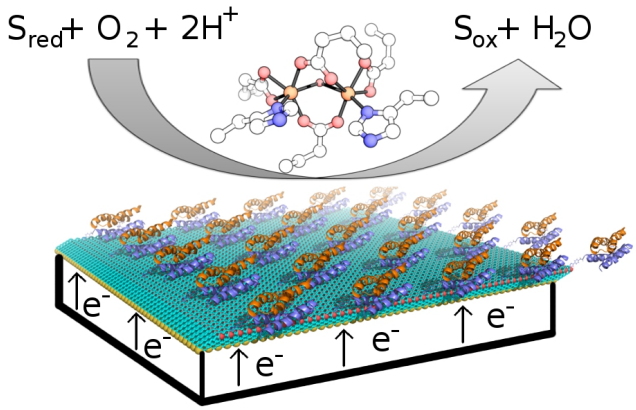


Figure 19. Pictorial representation of a biodevice able to convert orgainc substrates generating an amperometric response to measure their concentration in solution.

In perspective, this work is the first step in the ambitious achievement to obtain a low molecular weight synthetic protein able to monooxygenate organic compounds (Figure 19), with an outstanding impact in the field of environment pollutants analysis and degradation[118].

1.3.2. QF: design of a tetranuclear metal cofactor in a highly symmetric environment

Even though with the DF_{tet} subset there was the possibility to check for tetramerization of a single helix, a perfect D_2 symmetric compound was never tested for metal binding. With the QF (Quattro Ferro), we aim in the first *de novo* design of a tetranuclear metal cofactor bound in a symmetric coil. In this ambitious task, we were inspired by the cubane like Mn-O cluster found in Photosystem II, where four

manganese ions provide the oxidative equivalents necessary for the four electrons oxidation of water[119] (Figure 20).

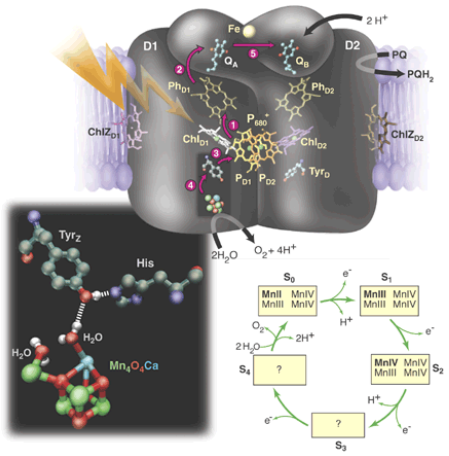


Figure 20. Photosystem II oxygen evolving complex. Adapted from reference [120].

Here, we report the design process leading to the core unit of the protein, synthesis and spectroscopic characterization of two analogues for solution study QF6 and QF33, as well as the design of QF in a crystal environment with the future aim to build up nanodevices for light harvesting.

References

(1) Dill, K. A.; Chan, H. S. *Nat. Struct. Mol. Biol.* **1997**, *4*, 10.
(2) Pauling, L.; Delbruck, M. *Science* **1940**, *92*, 77.
(3) Kolodny, R.; Pereyaslavets, L.; Samson, A. O.; Levitt, M. *Annu. Rev. Biophys.* **2013**, *42*, 559.
(4) Grant, A.; Lee, D.; Orengo, C. *Genome Biol.* **2004**, *5*, 107.
(5) Reardon, D.; Farber, G. K. *FASEB J.* **1995**, *9*, 497.
(6) Schneider, J. P.; Lombardi, A.; DeGrado, W. F. *Fold. Des.* **1998**, *3*, R29.
(7) Kamtekar, S.; Hecht, M. H. *FASEB J.* **1995**, *9*, 1013.

(8) Bork, P.; Holm, L.; Sander, C. *J. Mol. Biol.* **1994**, *242*, 309.
(9) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779.
(10) Lahr, S. J.; Engel, D. E.; Stayrook, S. E.; Maglio, O.; North, B.; Geremia, S.; Lombardi, A.; Degrado, W. F. *J. Mol. Biol.* **2005**, *346*, 1441.
(11) Muñoz, V.; Blanco, F. J.; Serrano, L. *Nat. Struct. Mol. Biol.* **1995**, *2*, 380.
(12) Maglio, O.; Nastri, F.; Calhoun, J. R.; Lahr, S.; Wade, H.; Pavone, V.; Degrado, W. F.; Lombardi, A. *JBIC J. Biol. Inorg. Chem.* **2005**, *10*, 539.
(13) Stanfield, R. L.; Wilson, I. A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 103.
(14) Fernandez-Fuentes, N.; Dybas, J. M.; Fiser, A. *PLoS Comput Biol* **2010**, *6*, e1000750.
(15) Chandonia, J.-M.; Brenner, S. E. *Science* **2006**, *311*, 347.
(16) Golovin, A.; Henrick, K. *BMC Bioinformatics* **2008**, *9*, 312.
(17) Johansson, M. U.; Zoete, V.; Michielin, O.; Guex, N. *BMC Bioinformatics* **2012**, *13*, 173.
(18) Johansson, M. U.; Zoete, V.; Guex, N. *Int. J. Mol. Sci.* **2013**, *14*, 7795.
(19) Schrödinger, L. L. C. The PyMOL Molecular Graphics System, Version 1.3r1, 2010.
(20) Lewit-Bentley, A.; Réty, S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 637.
(21) Stote, R. H.; Karplus, M. *Proteins Struct. Funct. Bioinforma.* **1995**, *23*, 12.
(22) Summa, C. M.; Lombardi, A.; Lewis, M.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **1999**, *9*, 500.
(23) Cowan, J. A. *Inorganic Biochemistry: An Introduction*; John Wiley & Sons, 1997.
(24) Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J. *Nature* **2009**, *460*, 823.
(25) Maglio, O.; Nastri, F.; Lombardi, A. In *Ionic Interactions in Natural and Synthetic Macromolecules*; Ciferri, A.; Perico, A.,

Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012; pp. 361–450.
(26) Lippard, S. J.; Berg, J. M. *Principles of bioinorganic chemistry*; University Science Books: Mill Valley, Calif., 1994.
(27) Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. *Struct. Lond. Engl. 1993* **1996**, *4*, 1171.
(28) Elrod-Erickson, M.; Benson, T. E.; Pabo, C. O. *Structure* **1998**, *6*, 451.
(29) Yadav, M. K.; Leman, L. J.; Price, D. J.; Brooks, C. L.; Stout, C. D.; Ghadiri, M. R. *Biochemistry (Mosc.)* **2006**, *45*, 4463.
(30) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401.
(31) Harbury, P. B.; Kim, P. S.; Alber, T. *Nature* **1994**, *371*, 80.
(32) Deng, Y.; Liu, J.; Zheng, Q.; Eliezer, D.; Kallenbach, N. R.; Lu, M. *Structure* **2006**, *14*, 247.
(33) Summa, C. M.; Rosenblatt, M. M.; Hong, J.-K.; Lear, J. D.; DeGrado, W. F. *J. Mol. Biol.* **2002**, *321*, 923.
(34) Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. *J. Mol. Biol.* **1987**, *196*, 657.
(35) Friedman, A. M.; Fischmann, T. O.; Steitz, T. A. *Science* **1995**, *268*, 1721.
(36) Samish, I.; MacDermaid, C. M.; Perez-Aguilar, J. M.; Saven, J. G. *Annu. Rev. Phys. Chem.* **2011**, *62*, 129.
(37) Butterfoss, G. L.; Kuhlman, B. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 49.
(38) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364.
(39) Li, Z.; Yang, Y.; Zhan, J.; Dai, L.; Zhou, Y. *Annu. Rev. Biophys.* **2013**, *42*, 315.
(40) Desmet, J.; Maeyer, M. D.; Hazes, B.; Lasters, I. *Nature* **1992**, *356*, 539.
(41) Koehl, P.; Delarue, M. *J. Mol. Biol.* **1994**, *239*, 249.
(42) Kono, H.; Saven, J. G. *J. Mol. Biol.* **2001**, *306*, 607.
(43) Voigt, C. A.; Gordon, D. B.; Mayo, S. L. *J. Mol. Biol.* **2000**, *299*, 789.

(44) Looger, L. L.; Hellinga, H. W. *J. Mol. Biol.* **2001**, *307*, 429.
(45) Calhoun, J. R.; Kono, H.; Lahr, S.; Wang, W.; DeGrado, W. F.; Saven, J. G. *J. Mol. Biol.* **2003**, *334*, 1101.
(46) Marchler-Bauer, A.; Lu, S.; Anderson, J. B.; Chitsaz, F.; Derbyshire, M. K.; DeWeese-Scott, C.; Fong, J. H.; Geer, L. Y.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Jackson, J. D.; Ke, Z.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Mullokandov, M.; Omelchenko, M. V.; Robertson, C. L.; Song, J. S.; Thanki, N.; Yamashita, R. A.; Zhang, D.; Zhang, N.; Zheng, C.; Bryant, S. H. *Nucleic Acids Res.* **2011**, *39*, D225.
(47) Hu, C.; Koehl, P. *Proteins* **2010**, *78*, 1736.
(48) Ho, B. K.; Curmi, P. M. G. *J. Mol. Biol.* **2002**, *317*, 291.
(49) Engel, D. E.; DeGrado, W. F. *Proteins* **2005**, *61*, 325.
(50) Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. *Nature* **2012**, *491*, 222.
(51) Zhang, J.; Grigoryan, G. In *Methods in Enzymology*; Elsevier, 2013; Vol. 523, pp. 21–40.
(52) Helling, R.; Li, H.; Mélin, R.; Miller, J.; Wingreen, N.; Zeng, C.; Tang, C. *J. Mol. Graph. Model.* **2001**, *19*, 157.
(53) Grigoryan, G.; DeGrado, W. F. *J. Mol. Biol.* **2011**, *405*, 1079.
(54) Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F. *Proc. Natl. Acad. Sci.* **2000**, *97*, 6298.
(55) Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G. *Nature* **1991**, *349*, 541.
(56) Frolow, F.; Kalb, A. J.; Yariv, J. *Nat. Struct. Biol.* **1994**, *1*, 453.
(57) deMaré, F.; Kurtz, D. M., Jr; Nordlund, P. *Nat. Struct. Biol.* **1996**, *3*, 539.
(58) Andersson, M. E.; Högbom, M.; Rinaldo-Matthis, A.; Blodig, W.; Liang, Y.; Persson, B.-O.; Sjöberg, B.-M.; Su, X.-D.; Nordlund, P. *Biochemistry (Mosc.)* **2004**, *43*, 7966.
(59) Lindqvist, Y.; Huang, W.; Schneider, G.; Shanklin, J. *EMBO J.* **1996**, *15*, 4081.

(60) Rosenzweig, A. C.; Brandstetter, H.; Whittington, D. A.; Nordlund, P.; Lippard, S. J.; Frederick, C. A. *Proteins* **1997**, *29*, 141.
(61) Sazinsky, M. H.; Lippard, S. J. *Acc. Chem. Res.* **2006**, *39*, 558.
(62) Stenkamp, R. E. *Chem. Rev.* **1994**, *94*, 715.
(63) Nordlund, P.; Sjöberg, B. M.; Eklund, H. *Nature* **1990**, *345*, 593.
(64) Broadwater, J. A.; Ai, J.; Loehr, T. M.; Sanders-Loehr, J.; Fox, B. G. *Biochemistry (Mosc.)* **1998**, *37*, 14664.
(65) Sigel, A.; Sigel, H. *Met.-Based Drugs* **1998**, *5*, 262.
(66) Berthold, D. A.; Andersson, M. E.; Nordlund, P. *Biochim. Biophys. Acta* **2000**, *1460*, 241.
(67) Bailey, L. J.; Fox, B. G. *Biochemistry (Mosc.)* **2009**, *48*, 8932.
(68) Benini, S.; Rypniewski, W. R.; Wilson, K. S.; Ciurli, S. *J. Inorg. Biochem.* **2008**, *102*, 1322.
(69) Bertini, I. *Biological Inorganic Chemistry: Structure and Reactivity*; University Science Books, 2007.
(70) Wilkinson, K. W.; Ford, G. C.; Moir, A. J.; Rice, D. W.; Rodgers, H. F.; Smith, J. M.; Stillman, T. J.; Goward, C. R. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, *53*, 197.
(71) Faraone-Mennella, J.; Gray, H. B.; Winkler, J. R. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6315.
(72) Crofts, A. R. *Annu. Rev. Physiol.* **2004**, *66*, 689.
(73) Cafaro, V.; Scognamiglio, R.; Viggiani, A.; Izzo, V.; Passaro, I.; Notomista, E.; Piaz, F. D.; Amoresano, A.; Casbarra, A.; Pucci, P.; Di Donato, A. *Eur. J. Biochem.* **2002**, *269*, 5689.
(74) Merkx, M.; Kopp, D. A.; Sazinsky, M. H.; Blazyk, J. L.; Müller, J.; Lippard, S. J. *Angew. Chem. Int. Ed.* **2001**, *40*, 2782.
(75) Culpepper, M. A.; Rosenzweig, A. C. *Crit. Rev. Biochem. Mol. Biol.* **2012**, *47*, 483.
(76) Bailey, L. J.; Acheson, J. F.; McCoy, J. G.; Elsen, N. L.; Phillips, G. N.; Fox, B. G. *Biochemistry (Mosc.)* **2012**, *51*, 1101.
(77) Cafaro, V.; Izzo, V.; Scognamiglio, R.; Notomista, E.; Capasso, P.; Casbarra, A.; Pucci, P.; Di Donato, A. *Appl. Environ. Microbiol.* **2004**, *70*, 2211.

(78) Sazinsky, M. H.; Bard, J.; Di Donato, A.; Lippard, S. J. *J. Biol. Chem.* **2004**, *279*, 30600.
(79) Bailey, L. J.; McCoy, J. G.; Phillips, G. N.; Fox, B. G. *Proc Natl Acad Sci U S A* **2008**, *105*, 19194.
(80) Mitchell, K. H.; Studts, J. M.; Fox, B. G. *Biochemistry (Mosc.)* **2002**, *41*, 3176.
(81) Tinberg, C. E.; Lippard, S. J. *Biochemistry (Mosc.)* **2010**, *49*, 7902.
(82) Lee, D.; Lippard, S. J. In *Comprehensive Coordination Chemistry II*; McCleverty, J. A.; Meyer, T. J., Eds.; Pergamon: Oxford, 2003; pp. 309–342.
(83) Lee, S. J.; McCormick, M. S.; Lippard, S. J.; Cho, U.-S. *Nature* **2013**, *494*, 380.
(84) Elsen, N. L.; Bailey, L. J.; Hauser, A. D.; Fox, B. G. *Biochemistry (Mosc.)* **2009**, *48*, 3838.
(85) Koder, R. L.; Dutton, P. L. *Dalton Trans.* **2006**, 3045.
(86) Nastri, F.; Bruni, R.; Maglio, O.; Lombardi, A. In *Coordination Chemistry in Protein Cages*; Ueno, T.; Watanabe, Y., Eds.; John Wiley & Sons, Inc., 2013; pp. 43–85.
(87) Handel, T.; DeGrado, W. F. *J. Am. Chem. Soc.* **1990**, *112*, 6710.
(88) Regan, L.; Clarke, N. D. *Biochemistry (Mosc.)* **1990**, *29*, 10878.
(89) Klemba, M.; Regan, L. *Biochemistry (Mosc.)* **1995**, *34*, 10094.
(90) Shiga, D.; Nakane, D.; Inomata, T.; Funahashi, Y.; Masuda, H.; Kikuchi, A.; Oda, M.; Noda, M.; Uchiyama, S.; Fukui, K.; Kanaori, K.; Tajima, K.; Takano, Y.; Nakamura, H.; Tanaka, T. *J. Am. Chem. Soc.* **2010**, *132*, 18191.
(91) Shiga, D.; Funahashi, Y.; Masuda, H.; Kikuchi, A.; Noda, M.; Uchiyama, S.; Fukui, K.; Kanaori, K.; Tajima, K.; Takano, Y.; Nakamura, H.; Kamei, M.; Tanaka, T. *Biochemistry (Mosc.)* **2012**, *51*, 7901.
(92) Xie, F.; Sutherland, D. E. K.; Stillman, M. J.; Ogawa, M. Y. *J. Inorg. Biochem.* **2010**, *104*, 261.
(93) Grzyb, J.; Xu, F.; Weiner, L.; Reijerse, E. J.; Lubitz, W.; Nanda, V.; Noy, D. *Biochim. Biophys. Acta BBA - Bioenerg.* **2010**, *1797*, 406.

(94) Choma, C. T.; Lear, J. D.; Nelson, M. J.; Dutton, P. L.; Robertson, D. E.; DeGrado, W. F. *J. Am. Chem. Soc.* **1994**, *116*, 856.
(95) Robertson, D. E.; Farid, R. S.; Moser, C. C.; Urbauer, J. L.; Mulholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; DeGrado, W. F.; Dutton, P. L. *Nature* **1994**, *368*, 425.
(96) Koder, R. L.; Valentine, K. G.; Cerda, J.; Noy, D.; Smith, K. M.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **2006**, *128*, 14450.
(97) Koder, R. L.; Anderson, J. L. R.; Solomon, L. A.; Reddy, K. S.; Moser, C. C.; Dutton, P. L. *Nature* **2009**, *458*, 305.
(98) Ghirlanda, G.; Osyczka, A.; Liu, W.; Antolovich, M.; Smith, K. M.; Dutton, P. L.; Wand, A. J.; DeGrado, W. F. *J. Am. Chem. Soc.* **2004**, *126*, 8141.
(99) Bender, G. M.; Lehmann, A.; Zou, H.; Cheng, H.; Fry, H. C.; Engel, D.; Therien, M. J.; Blasie, J. K.; Roder, H.; Saven, J. G.; DeGrado, W. F. *J. Am. Chem. Soc.* **2007**, *129*, 10732.
(100) Cochran, F. V.; Wu, S. P.; Wang, W.; Nanda, V.; Saven, J. G.; Therien, M. J.; DeGrado, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 1346.
(101) Fry, H. C.; Lehmann, A.; Saven, J. G.; DeGrado, W. F.; Therien, M. J. *J. Am. Chem. Soc.* **2010**, *132*, 3997.
(102) Geremia, S.; Di Costanzo, L.; Randaccio, L.; Engel, D. E.; Lombardi, A.; Nastri, F.; DeGrado, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 17266.
(103) Maglio, O.; Nastri, F.; Pavone, V.; Lombardi, A.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3772.
(104) Pasternak, A.; Kaplan, J.; Lear, J. D.; Degrado, W. F. *Protein Sci.* **2001**, *10*, 958.
(105) Marsh, E. N. G.; DeGrado, W. F. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5150.
(106) Kaplan, J.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11566.
(107) Faiella, M.; Andreozzi, C.; de Rosales, R. T. M.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A. *Nat Chem Biol* **2009**, *5*, 882.

(108) Maglio, O.; Nastri, F.; Martin de Rosales, R. T.; Faiella, M.; Pavone, V.; DeGrado, W. F.; Lombardi, A. *Comptes Rendus Chim.* **2007**, *10*, 703.
(109) Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; Degrado, W. F. *Nat. Chem.* **2012**, *4*, 900.
(110) Bhushan, B. *Philos. Transact. A Math. Phys. Eng. Sci.* **2009**, *367*, 1445.
(111) Winkler, J. R.; Gray, H. B. *Chem. Rev.* **2013**.
(112) Sanghamitra, N. J. M.; Ueno, T. *Chem. Commun.* **2013**, *49*, 4114.
(113) Banwell, E. F.; Abelardo, E. S.; Adams, D. J.; Birchall, M. A.; Corrigan, A.; Donald, A. M.; Kirkland, M.; Serpell, L. C.; Butler, M. F.; Woolfson, D. N. *Nat. Mater.* **2009**, *8*, 596.
(114) Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N. *Science* **2013**, *340*, 595.
(115) Laganowsky, A.; Zhao, M.; Soriaga, A. B.; Sawaya, M. R.; Cascio, D.; Yeates, T. O. *Protein Sci. Publ. Protein Soc.* **2011**, *20*, 1876.
(116) Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; Degrado, W. F. *Science* **2011**, *332*, 1071.
(117) Lanci, C. J.; MacDermaid, C. M.; Kang, S.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. *Proc. Natl. Acad. Sci.* **2012**, *109*, 7304.
(118) Zhong, Z.; Fritzsche, M.; Pieper, S. B.; Wood, T. K.; Lear, K. L.; Dandy, D. S.; Reardon, K. F. *Biosens. Bioelectron.* **2011**, *26*, 2407.
(119) Lubitz, W.; Reijerse, E. J.; Messinger, J. *Energy Environ. Sci.* **2008**, *1*, 15.
(120) Rutherford, A. W.; Boussac, A. *Science* **2004**, *303*, 1782.

Chapter 2: *De novo* design of DF-Click

heterodimer, Click chemistry, DOSY-NMR, circular dichroism

2 *De novo* design of DF-Click

2.1 Results and discussion

In this work we developed a new approach for the synthesis of asymmetrical DF1 analogues, called DF-Click. In order to induce correct heterodimerization, a protocol involving side chain/main chain covalent binding via Click Chemistry was developed. Here, we want to proof the applicability of the method by the *de novo* design, synthesis and spectroscopic characterization of the first compound of this series, named DF-Click1. The analysis of DF-Click1 supported the correctness of the full design process. These achievements are crucial in the generation of a library of DF-based metal ligands for the screening of new function/substrates.

In previously synthesized asymmetric DF analogues, asymmetry was achieved in a context of a non-covalent self-assembling heterotetramer[1] or of a long-chain monomer without any internal pseudo-symmetry[2]. DF_{tet} subset demonstrated to be useful in the screening of new compounds, but the structural characterization of its members was limited by the difficult estimation of the desired tetramer

in solution[1,3,4]. DF_{SC}, on the other side, was amenable to structural characterization but, as proven by the 3His-DF_{SC} case, a rigorous and time-consuming design process was necessary to obtain a stable compound[2], thus precluding its applicability as scaffold for catalytic screening purposes. These assumptions moved our attention in the exploitation of the wide-studied helix-loop-helix formula adopted in DF1. Helix-loop-helix dimerization guarantees a robust structural framework, as well as an easy-to-screen system. For example, one can keep fixed only one of the two subunits, while screening the second and *vice versa*.

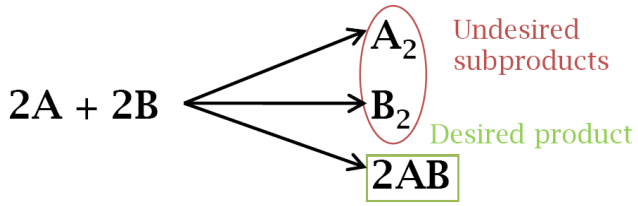


Figure 2.1. Simple mixing of monomeric subunits may lead to the formation of undesired homodimeric products.

An intrinsic difficulty in the synthesis of heterodimers is the undesired formation of homodimeric byproducts (Figure 2.1). Homodimerization propensity must be suppressed and, concurrently, heterodimerization enhanced. There are basically three possible strategies:

- Covalent binding: by means of a linker or through side-chain modifications.
- Assembly on solid support: by means of surface immobilization.

- Self-Assembly: by means of molecular recognition and weak interactions.

Here, we report our results achieved through the first strategy, and we demonstrate that a rational design is indeed necessary for the success of the synthetic protocol. In particular, NCAAs (non-canonical amino acids) were adopted in our protocol in order to accomplish the side chain modification strategy. A very efficient and orthogonal chemistry was crucial to selectively bind the two different monomers.

In 2001 Kolb, Finn and Sharpless published their seminal paper, describing few good reactions providing diverse chemical functionality, and coined the term "Click Chemistry"[5]. "Click Chemistry" has found diverse applications, from drug synthesis to polymer chemistry. Taking guidance from nature way of creating biomolecules and keeping in sight a "process chemistry" approach, Sharpless *et al.* described a set of stringent criteria to qualify a reaction as "click chemistry". This reaction must be stereospecific, but not necessarily enantioselective, wide in scope, modular, must give very high yields, and should produce only inoffensive byproducts. It is important to recognize that click reactions achieve their required features by having a high thermodynamic driving force, usually greater than 20 kcal mol^{-1}. Such processes proceed smoothly to completion, and tend to be highly selective for a single product. Therefore, Sharpless et al presented click chemistry as a set of reactions that can be envisioned to be "spring-loaded" for a single trajectory.

Chapter 2: De novo design of DF-Click

One of the most studied among the non-biological linkages is the triazole ring formed *via* 1,3-dipolar cycloaddition of an organic azide to a terminal alkyne. The Huisgen 1,3-dipolar cycloaddition, was found to be the most useful type of reaction among all the click reactions. However, the traditional Huisgen cycloaddition reaction of organic azides with terminal alkynes doesn't meet the criteria of a click reaction, because it requires high temperature, often more than 100 °C, for several hours and it produces a mixture of 1,4- and 1,5-triazoles, without regiospecificity (Figure 2.2)[6] .

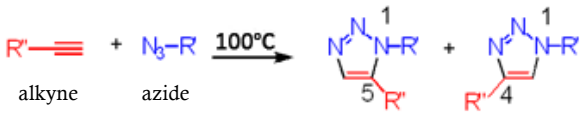


Figure 2.2. Azide alkyne Huisgen 1,3 dipolar cycloaddition.

The Huisgen 1,3-dipolar cycloaddition reaction regained attention in recent years, due to the introduction in 2001 of Cu(I) catalysis by Tornøe and Meldal[7], leading to a major improvement in both rate and regioselectivity of the reaction.

Since the discovery that Cu(I) dramatically enhances the rate and regioselectivity of the reaction, this catalyzed reaction has been known as CuAAC (Cu(I)-catalyzed Azide-Alkyne 1,3-dipolar Cycloaddition).

The mechanistic role of the copper ion, in increasing by seven orders of magnitude, the reaction rate has been subject of many disputes and revisions. Moreover, this reaction is 1,4-regioselective, as shown in Figure 2.3.

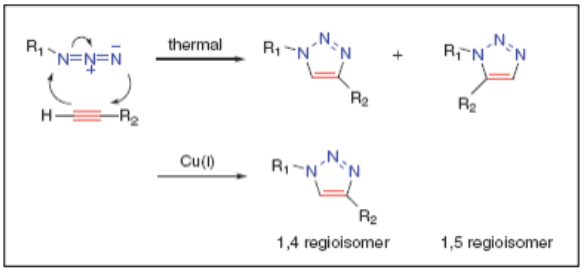


Figure 2.3. CuAAC regioselectivity: only the 1,4 disubstituted triazole is obtained.

On the basis of thermal and kinetic studies, the CuAAC process has been proposed to occur via a stepwise mechanism as shown in Figure 2.4[7].

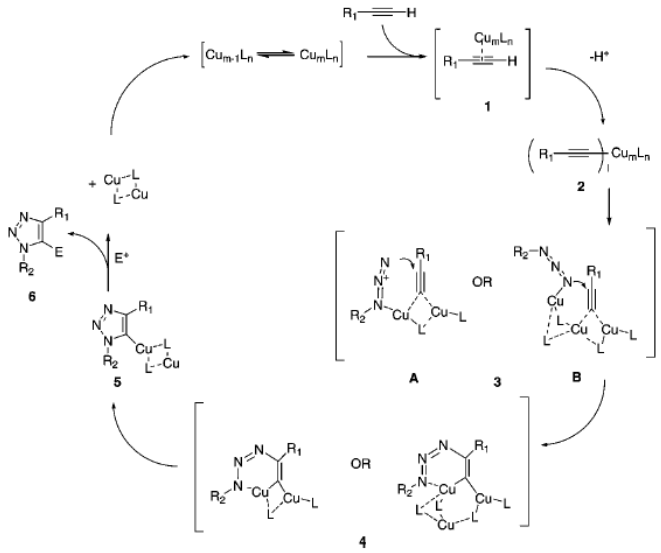


Figure 2.4. CuAAC proposed mechanism.

The reaction is of the second order both with respect to [Cu(I)] and [alkyne] [7].

The first step in the catalytic cycle is the formation of Cu(I)-acetylide complex, by coordination of Cu^{+} with acetylene π-electrons. The π-coordination of Cu(I) lowers the pK_a of the acetylene proton,

thus allowing the exothermal formation of the acetylide complex. Subsequently, the Cu^{+}-acetylide complex coordinates the azide and then the activated azide-Cu(I)-acetylide specie cyclizes to form the copper metallocycle. Then, this intermediate undergoes facile ring contraction, to give a copper-triazole complex. Finally, this last compound eventually releases the free triazole and the $L_nCu(I)$, by protonation or reaction with other electrophiles[7].

Many research groups published reviews and papers about the different applications of the Click chemistry in numerous fields, such as pharmaceutical science[8], drug discovery and synthesis[9], and polymer bioconjugation[10]. These various applications of click chemistry, and in particular the use of Cu(I) catalyzed azide-alkyne cycloaddition, prompted us to test this reaction in covalent binding of two functionalized peptides, to create heterodimeric proteins.

In the worst-case scenario, starting monomers could be both engaged in homodimeric interactions. In such a case, entropic driving force given by the covalent binding may not be enough to prevent the formation of high molecular weight oligomers composed by several alternating homodimers (Figure 2.5).

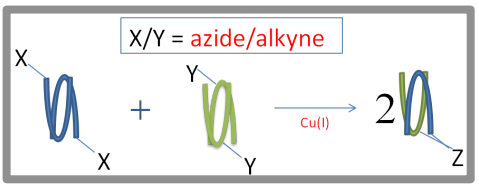


Figure 2.5. Two helix-loop-helix homodimers react to give only the desired product, if it is preferred to high oligomeric states.

In such a case, careful rational design of heterodimeric interaction is particularly relevant. In proteins, polar interactions contribute to global energy of folding, only in a limited extent[11]. However, it has been shown that H-bond between side chains in proteins core improve the folding selectivity, by increasing the energy of alternative misfolded states[12,13]. Here, we show how to design such interactions in the DF scaffold, taking inspiration from the core of hydroxylase components of BMMs and by the successful design of 3His-DF_{SC}.

2.1.1 Design

Design of 'Clickable' subunits

Aimed to design a covalent heterodimer, the amino acid positions suitable for the Click reaction have to be selected. Mutation of residues, which could be important in terms of stability of the folding or reactivity, were carefully avoided, still permitting a ligation, which could effectively help subunits recognition. Further on, for keeping the synthetic procedures easy and inexpensive, therefore 6-azido-hexanoic acid (6-aha) and propargyl glycine (Pra) were selected, as the azide and alkyne moieties, respectively. Both moieties are widely adopted in the literature, e.g. for the introduction of exogenous probes in proteins, or to make cyclic peptides [14–17].

Structural study in solution by NMR of DF3, the most active member of DF family[18], shows a plasticity of the helix 1 N-termini, as shown in Figure 2.6. The NMR structure of helices 1 and 1' in Zn(II)-DF3 present a high flexibility, which may be ascribed to the two Gly residues introduced to increase active site accessibility[18].

Hence, the N-termini was chosen as the azide position, just replacing acetic anidride with 6-aha, in the final capping step, keeping the synthesis as easy as possible. We define 'Z' this peptide subunit, bearing the azide moiety.

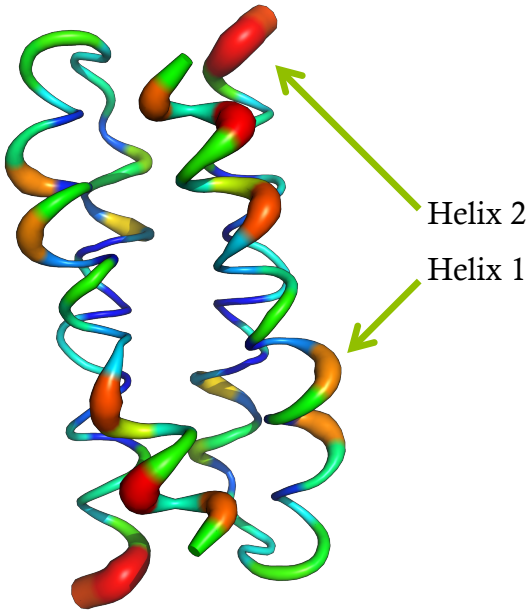


Figure 2.6. Worm-like representation of Zn-DF3 (PDB access code: 2kik). Red thicker residues show the highest displacement in the bundle, blue thinner show the lowest.

Therefore, computer-aided design was adopted to decide the best position for Pra mutation. For this purpose, the DF3 interchain distances between the N-termini capping acetyl carbon, in one chain, and the C-alpha atoms, in the other were measured (Figure 2.7, top).

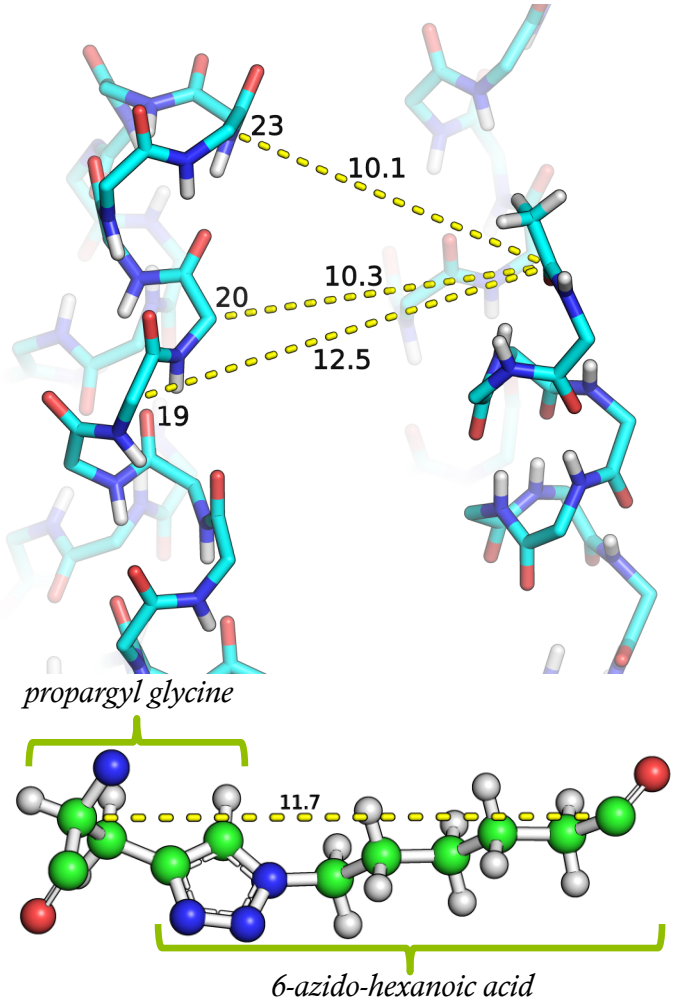


Figure 2.7. Acetyl carbon Cα distances as found in DF3 (top). Residues 19, 20 and 23 of helix 1 display the best agreement with calculated distance for the linker (bottom).

Since the distance crossed by the linker is 11.7 Å (Figure 2.7, bottom), those residues giving a 2 Å maximum deviation from this ideal value were the best candidates for positioning the Pra residue. Loop residues were excluded in this search, since loop conformation

exerts a crucial role in the stability of the protein[18,19] . Therefore, positions 19, 20 and 23 satisfied our requirement giving distances of 12.5, 10.3 and 10.1 Å, respectively.

The best fit of the linker against the selected residue positions was evaluated, while keeping free torsion of the linker bonds. Only one gauche conformation was needed, when 23 residue was set as the anchoring position. This analysis resulted in mutation of tyrosine 23 to propargylglycine. We define 'K' the subunit bearing the alkyne moiety.

Design of a Toluene Monooxygenase mimic: DF-Click1

As demonstrated by the several variants already synthesized[20], redesign of DF scaffold is simplified by its reliability versus several mutations, due to its structural stability. In Figure 2.8, the sequence alignments of sMMOH (soluble methane monooxygenase hydroxylase), ToMOH (toluene/o-xylene monooxygenase hydroxylase), T4moH (toluene 4-monooxygenase hydroxylase), and DF3 together with the new sequence of DF-Click1, derived from this redesign process, are reported for comparison.

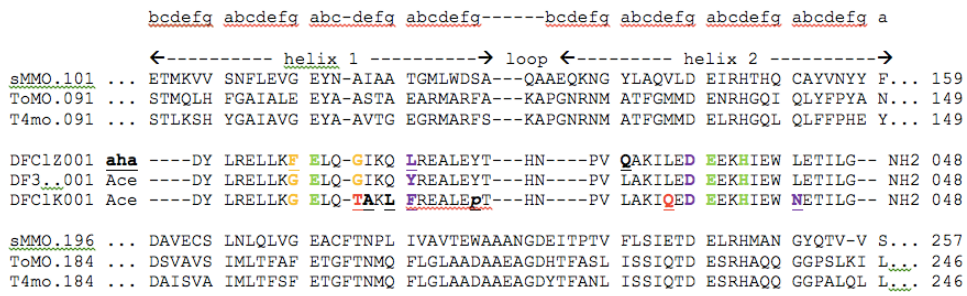


Figure 2.8. Structure based sequence alignment of some BMMs with respect to DF3 and the new heterodimer DF-Click1. The coordinating residues are highlighted in green, second-shell ligands in purple, residues involved in catalysis in red, and residues involved in active site access in orange. The new substitutions made in the DF3 sequence are underlined. Italic "p" stands for Propargyl Glycine. Residue numbering from PDB access code: 1MTY for sMMO, 2INC for ToMO, 3DHI for T4mo, 2KIK for DF3.

Some crucial aspects can be highlighted:

- **Active site access**: in BMMs substrate is delivered to the catalytic site through an internal channel, and a final Phe-rich pocket recognize and stabilize the substrate[21].
- **Proton transfer catalysis**: two protons are required during dioxygen activation. Some hydrogen-bond pattern is mandatory to accomplish hydroxylase function. Recently, has been proven the role of Thr 201 (213 in sMMO) and Gln 229 (242 in sMMO) in stabilizing a water molecule in the active site, involved into proton transfer[22–24].

Redesign was accomplished through successive energy minimization steps, starting from DF1-L13G crystallographic coordinates (PDB entry: 1LT1). Zinc(II) was adopted as the metal in this phase of the design, because its parameterization is particularly reliable. Each substitution was followed by a minimization cycle, in order to relax the structure. A final round of unconstrained

minimization was performed after the insertion of the linker. Mutations with respect to DF3 can be summarized in three classes:

- **Second-shell access and recognition substitutions**: Y17L, G9F for chain Z; Y17F for chain K.
- **Second-shell proton transfer substitution**: G13T, L33Q, L43N for chain K.
- **Structural substitutions**: L29Q for chain Z, I14A, Q16L for chain K.

Substitutions from the last class were inspired from molecular design and stability criterions. In particular, L29Q substitution for chain Z is required to better support L43N substitution, in order to create a hydrogen bond pattern throughout the protein (Figure 2.9C). Our aim is to improve the selectivity towards the heterodimerization; this H-bond configuration is strictly retained only in the heterodimer, while homodimer formation would give unsatisfied H-bond donors/acceptors in the hydrophobic core. I14A for chain K was chosen in order to open up a water gate toward the catalytic site, as found in ToMOH upon binding of the effector protein[24] (Figure 2.9B, Figure 2.9C). Q16L for chain K substitution was adopted to better accommodate the Phe 9 from chain Z. Though this small hydrophobic patch may help in the recognition of small organic compounds, a lower solubility was expected, with respect to DF3.

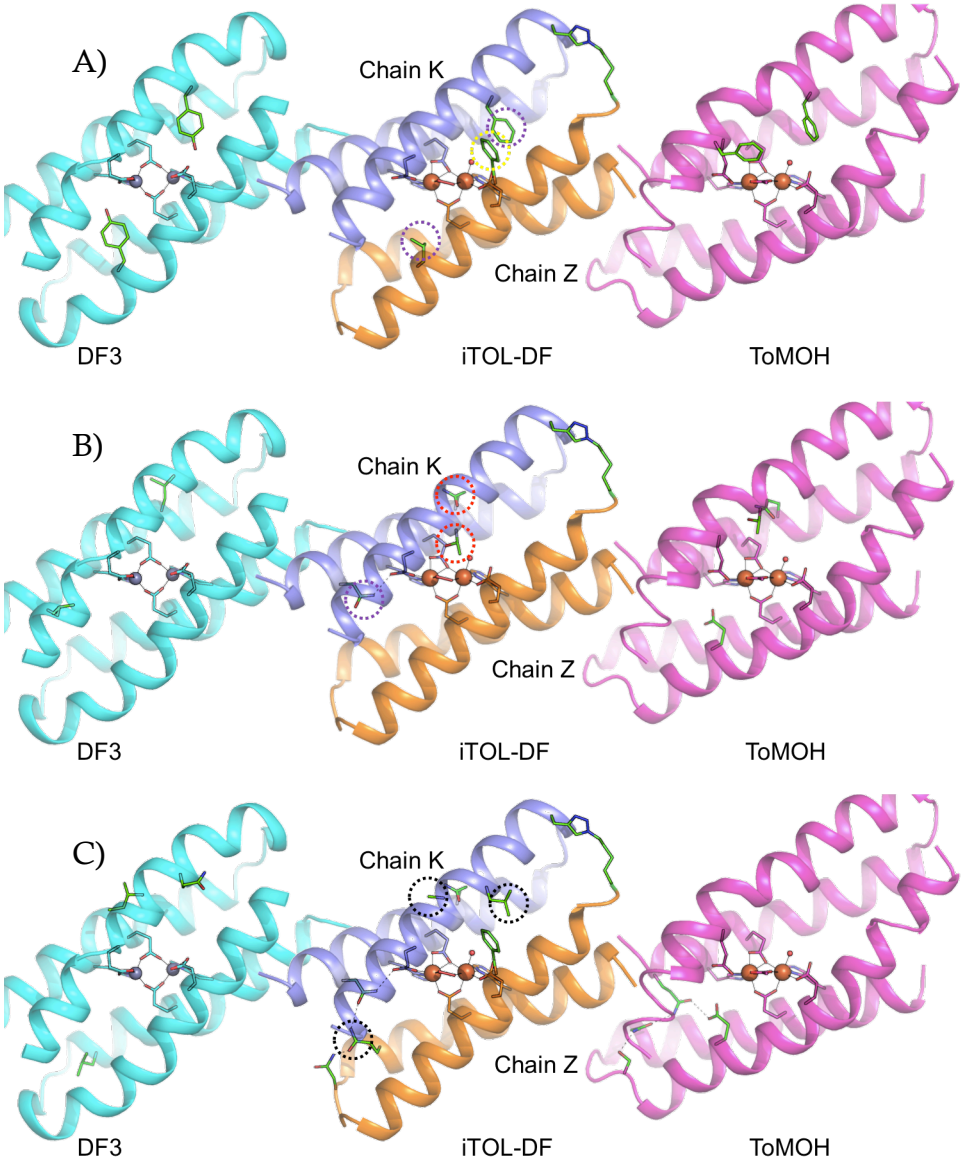


Figure 2.9. Mutations with respect to DF3 are compared to ToMOH. Mutations are grouped as in the text: A) Second-shell access and recognition substitutions, B) Second-shell proton transfer substitution, C) Structural substitutions. Circles highlight key residues, and the color code of **Figure 2.8** is retained.

2.1.2 Synthesis

Synthesis and Click reaction

Chain Z and chain K were chemically synthesized by standard Fmoc-protocols of solid-phase peptide synthesis[25,26], and purified *via* HPLC to a homogeneity. Successful heterodimer formation was achieved in 3.5 hours by addition of 2.5 eq of CuSO4 and 5.0 eq of tris(2-carboxyethyl)phosphine (TCEP) in buffered solution at pH 8.0. Reaction was followed by analytical HPLC (Figure 2.10) and product formation ascertained by ESI-MS.

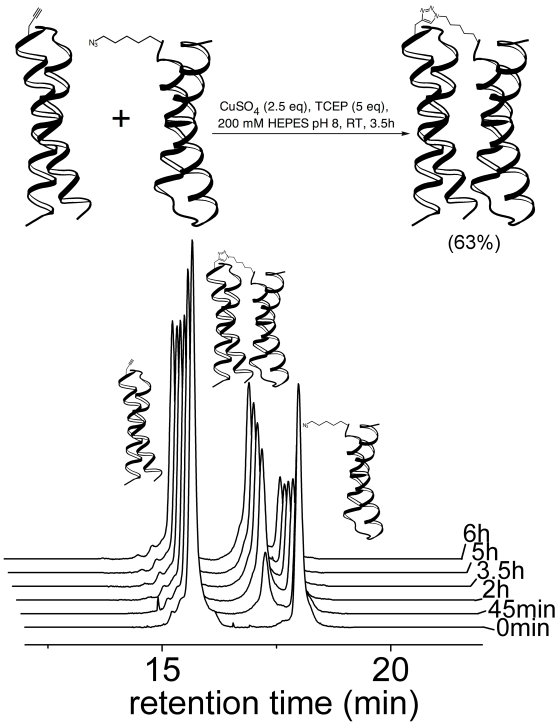


Figure 2.10. DF-Click1 reaction scheme (top), reaction progress has been followed by analytical HPLC (bottom).

Product purification was performed by preparative HPLC, with a final yield of 63%, based on the theoretical 1:1 monomer ligation.

As we developed a protocol in non-denaturing conditions, we tested the hypothesis that CuAAC could be adopted to test the goodness of the design process. We deliberately synthesized the two subunits with swapped positions in 9 and 13 (F9G, G13F for Z; G9T, T13G for K). Such mutations should destabilize the four-helix bundle, as we can predict by design (Figure 2.11).

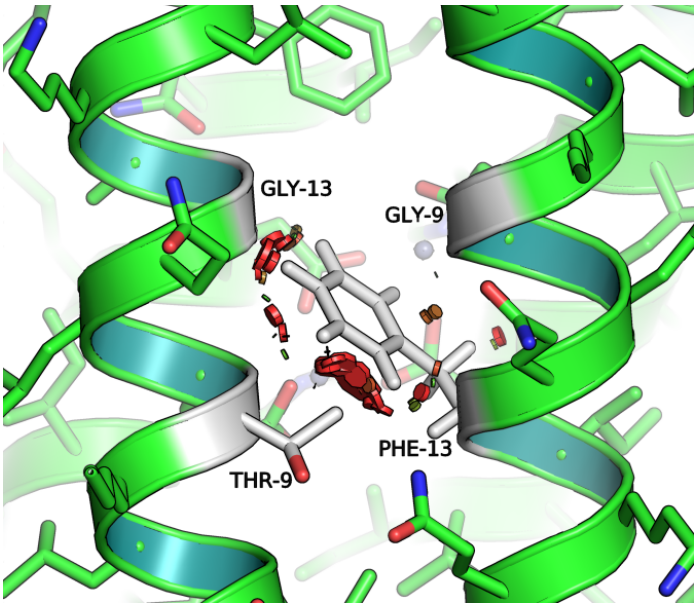


Figure 2.11. Swapped dimer (F9G, G13F for Z, G9T, T13G for K). Steric clashes are depicted as red circles. Best rotamer was adopted for Phe 13.

By applying our protocol of chemical ligation, through CuAAC, to this system the yield was less than 5% in 3.5 hours (Figure 2.12), thus demonstrating the advantage of Click chemistry as a critical test to the

design process. This aspect is of particular interest from the combinatorial viewpoint.

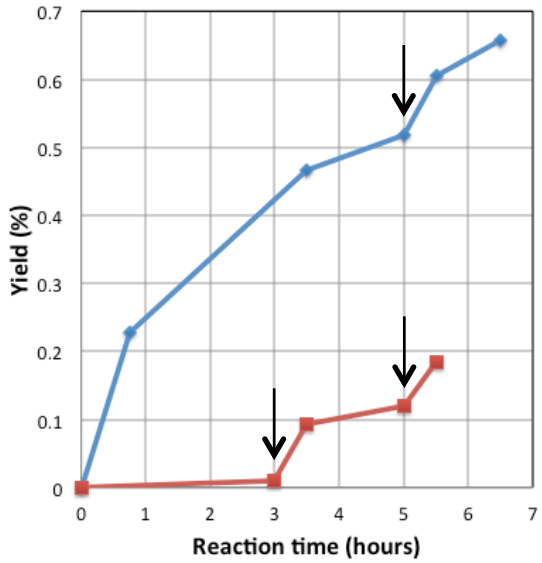


Figure 2.12. Yield as a function of reaction time as calculated by HPLC peak area. DF-Click1 (blue diamonds) and swapped DF-Click1 (red squares) reactions are compared. Black arrows indicate a further addition of catalyst.

Indeed, Click Chemistry may be adopted to select best combination of monomeric subunits in a pool of possible candidates. We are working on the exploitation of this finding, by developing a functional screening, where only dimers showing activity are selectively 'clicked', and hence purified, among several mixed monomers.

2.1.3 Spectroscopic characterization

2.1.3.a NMR and DOSY-NMR spectroscopies

Spectroscopic characterization: NMR

The DF-Click1 ability to bind Zn ion, in a defined globular folding, was tested by NMR spectroscopy. Apo-DF-Click1 shows only a slight structured content in 50 mM NaP buffer at pH 6.5. A finer dispersion of chemical shifts is observed upon $ZnCl_2$ addition: narrow and well-dispersed line-widths in the amide and aliphatic regions clearly suggest a well-defined globular folding (Figure 2.13). In particular, several resonances, shifted in the region -0.5 – 1.0 ppm, can be attributed to aliphatic side chains sequestered in the core of the protein[27]. Furthermore, the presence of two peaks around 15 ppm, assigned to histidine Nε protons, clearly confirms non-equivalent zinc binding histidines. This finding supports the hypothesis of an asymmetry in the metal binding site, as intended in the design. NOE spectra (data not shown) present several peaks compatible with a 96-residues protein, however peak assignment is difficult without isotope-labeled residues, and hence further studies are needed to completely solve the tridimensional structure.

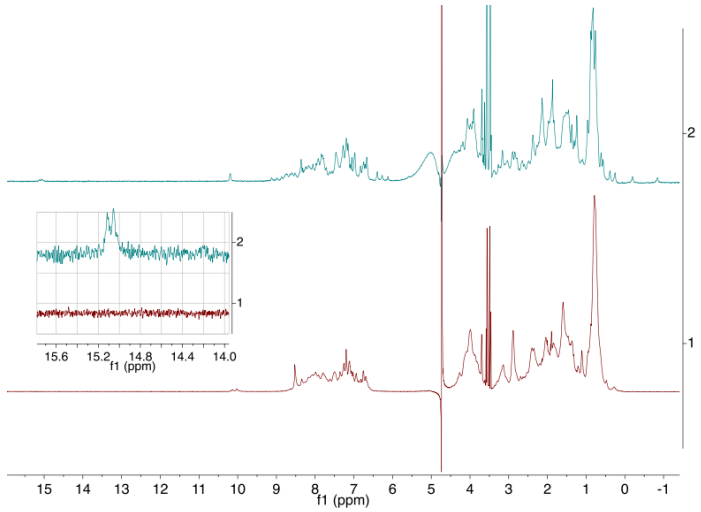


Figure 2.13. [1]H-NMR spectra of apo (red) and zinc bound (blue) DF-Click1 are compared. Inlet shows N^{ε} proton signals demonstrating nonequivalent binding histidines.

To further test the correctness of the design, the hydrodynamic radius (R_H) of the Zn-DF-Click1 by means of DOSY-NMR technique[28] was evaluated. The protein appears as a unique species, with an R_H of 24±2 Å, in good agreement with the theoretical calculated radius for the model (R_H^{teor}=23.4 Å) (Figure 2.14)[29]. This finding confirms the intended heterodimeric state, without the formation of higher order oligomers.

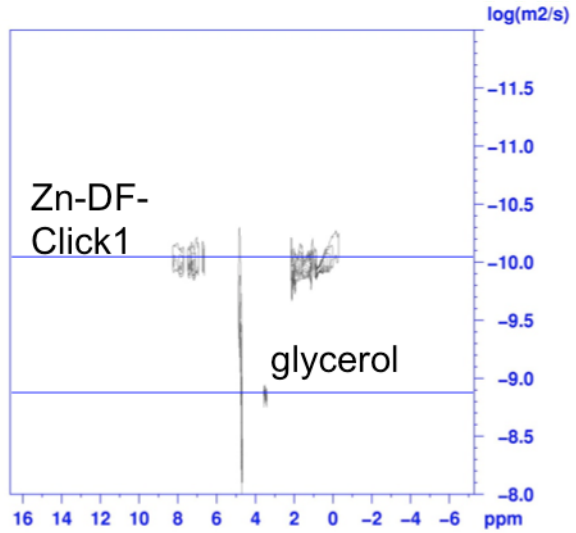


Figure 2.14. Bidimensional DOSY spectrum, x axis is the chemical shift, while y axis is the translational diffusion coefficient.

2.1.3.b CD spectroscopy

To demonstrate helical global folding of DF-Click1 circular dichroism spectroscopy was used.

Circular dichroism (CD) analysis of the two separate Z and K chains shows that only Z has a good α-helical content at pH 7.0 (Figure 2.15, left). Interestingly, the sum of the two spectra of each subunit is substantially different from the spectrum of the Click product DF-Click1. This feature shows that a positive interaction between the two conjugated chains occurs, bringing to a globally better folded structure (Figure 2.15, right), as found in heterotetramers from the DF_{tet} subset[1].

Further increase of mean residue ellipticity (m.r.e.) at 222 nm was observed upon addition of 10% excess $ZnCl_2$ (Figure 2.16). This finding indicates that DF-Click1 binding of Zn(II) induces an improvement of the global folding, as evinced from NMR analysis.

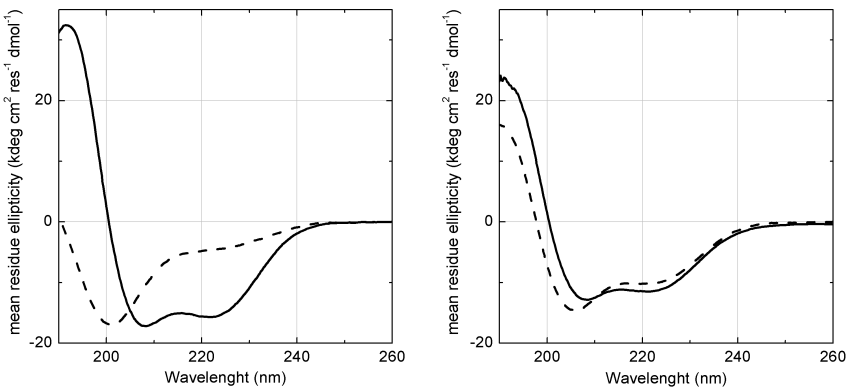


Figure 2.15. Circular dichroism spectra of Z (solid) and K (dashed) chains are superimposed on the left. On the right CD spectrum of the apo DF-Click1 (solid) is compared to the spectrum obtained by the sum of the spectra of the monomers (dashed).

Such a dramatic increase in structured content is expected at higher pH values, where the cost of ceiling negative charged glutamates in the hydrophobic core is too high in the absence of the positively charged metal ions[30,31] .

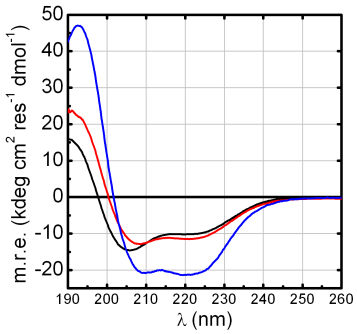


Figure 2.16. Zn-DF-Click1 CD spectrum (blue) is distinctive of a well-folded α protein, with a maximum at 192 nm and two minima at 208 and 222 nm. Apo-DF-Click1 spectrum (red) and monomers sum spectrum (black) are superimposed for comparison.

It is possible to follow the increase of helical content upon addition of several metal ions, to test metal binding properties of the new compound. Minimum at 222 nm decreases smoothly upon addition of Zn^{II}, Fe^{II}, Co^{II}, Cu^{II}, resembling a saturation mechanism of binding. Thus, the CD titration data were fit to a general $P+nM \leftrightarrows M_nP$ equilibrium (Figure 2.17).

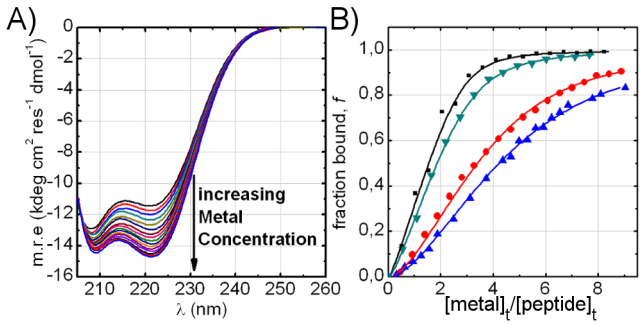


Figure 2.17. (A) $CoCl_3$ titration into 8 μM DF-Click1 in 10 mM phosphate buffer pH 7.0 monitored by circular dichroism. (B) Plot of normalized bound peptide (theta 222) as a function of Zn(II) (black squares), Fe(II) (dark cyan down triangles), Co(II) (red circles), Cu(II) (blue triangles) metal: peptide concentration ratio. Solid lines represent the best fit with the binding equation reported in the experimental section.

For all metal ions, the expected protein:metal ratio of 1:2 was observed, with K_D values are listed in Table 2-1. Interestingly, no metal binding was detected for Mn^{II} and Ni^{II}, while a significant selectivity towards Zn^{II} and Fe^{II} was observed. Deviation from the Irving-Williams series was already observed for DF compounds[32], but this selectivity was never observed in the other members of the DF-series[19], suggesting that the metal binding site is rigid enough to impart a predefined geometry to the metal centers. Nevertheless, it is difficult to explain the subtle interplay of factors that may affect the selectivity towards a single metal, as in the case of MnFe RNR[33], or CuZn SOD[34]. A deeper spectroscopic analysis may be necessary to understand the role played by the asymmetric context in driving the metal binding selectivity. The results available to date demonstrated that the small DF scaffold might be finely tuned to reproduce metal binding behavior of natural proteins. This was made possible by designing asymmetrical first and second coordination shells, demonstrating the relevance of the new approach adopted in this work.

Table 2-1. Metal binding parameters as evaluated from CD titration experiments.

Metal	**Metal: Protein ratio, n**	**Dissociation Constant, K_D (pM^2)**
Zn^{2+}	2.0 ± 0.1	8 ± 1
Fe^{2+}	2.00 ± 0.04	5.9 ± 0.2
Co^{2+}	2.06 ± 0.04	216 ± 3
Cu^{2+}	2.01 ± 0.04	222 ± 2
Mn^{2+}	ND	ND
Ni^{2+}	ND	ND

2.1.3.c UV-Visible absorption spectroscopy

Co(II) was selected for preliminary study on the metal geometry, because cobalt complexes show distinctive UV-Vis spectra, which are sensitive to the metal coordination geometry and to the nature of the ligands[35]. Figure 2.18 shows Co(II)-DF-Click1 absorption bands at wavelengths of 519, 548, and 581 nm. The observed molar extinction coefficients per Co(II) ion are 51 $M^{-1}cm^{-1}$, 56 $M^{-1}cm^{-1}$, and 41 $M^{-1}cm^{-1}$, respectively (see Table 2-2), and they are distinctive of a pentacoordinate Co(II) complex.

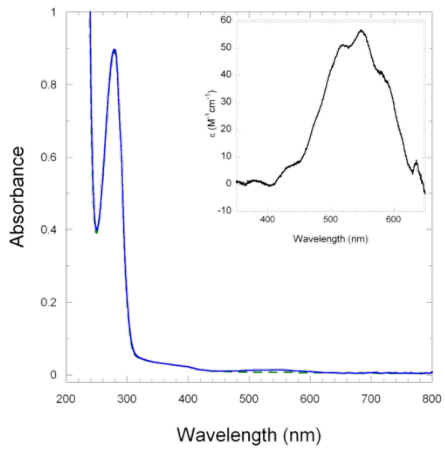


Figure 2.18. UV-Visible spectra of apo-DF-Click1 (green, dashed) and Co(II)-DF-Click1 (blue, solid) are superimposed for comparison. Inlet shows a zoom of the d-d transition band of the Co(II) complex.

Table 2-2. Spectroscopic data for various Co(II) substituted dinuclear metalloproteins

Protein	**λ (nm)**	**ε ($M^{-1}cm^{-1}$)**	**geometry**
DF-Click1	519; 548; 581	51; 56; 41	penta-
DF3[19]	522; 548; 598	88; 98; 66	penta-
MMOH[23]	520	40	dist. ott-
R2-RNR[35]	515; 550	N.D.; 115	penta-

Chapter 2: De novo design of DF-Click

DF-Click1 efficiently binds iron(III). The complex was obtained upon addition of stoichiometric amount of Fe(II) to a pH 7.0 solution of the apo protein, under Ar atmosphere, and successive exposition to air for one hour. After several steps of centrifugation and microfiltration, a brown solution was obtained. The UV-Vis spectrum shows an intense shoulder at 350 nm, a typical charge transfer band indicative of the Fe^{III}-O-Fe^{III} bridge formation[37] (Figure 2.19). Furthermore, DF-Click1 does not present the spectral feature at 500 nm previously observed in other DF variants, where an unattended tyrosinate binding to the metal cofactor produces a LMCT band at those wavelengths.

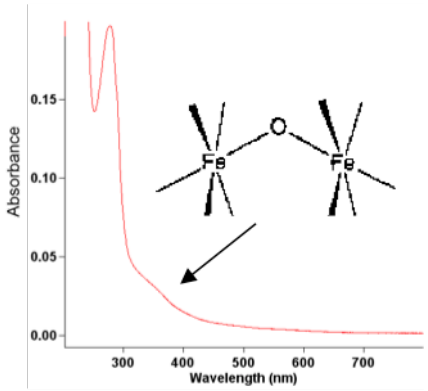


Figure 2.19. UV-Visible spectrum of the di-iron DF-Click1 complex, showing the oxo-bridge feature at 350 nm.

Table 2-3 DF-Click1 and DF3 oxo-bridge CT are tabulated for comparison.

Protein	λ (nm)	ε(M^{-1} cm^{-1})
DF-Click1	344	2638
DF3	350	5270

2.2 Experimental section

Peptide synthesis was carried out using reagent grade anhydrous solvents, whereas for the analysis and chromatographic purifications HPLC grade solvents were employed. Solvents with a higher degree of purity were used in the preparation of solutions for UV-Vis and CD investigations. All the solvents adopted in the peptide synthesis and purification, were provided by Romil. Piperidine and triisopropylsilane (TIS) were supplied by Sigma Aldrich; trifluoroacetic acid and N,N-diisopropyletyl amine were from Applied Biosystem and acetic anhydride was from Fluka. All reagents were used without further purification.

The H-PAL ChemMatrix resin (substitution level 0.31 mmol/g) was from Sigma Aldrich and all the N-α-Fmoc amino acids and the 6-azidohexanoic acid were purchased from Merck Millipore NovaBiochem; activating HOBt, HATU and HCTU were provided by LC Sciences, Anaspec and NovaBiochem, respectively.

$CuSO_4 \cdot 5H_2O$ and TCEP for CuAAC reaction were from Sigma Aldrich.

Zinc (II) chloride, cobalt (II) chloride, manganese (II) chloride, nickel (II) chloride, Cu (II) chloride and ammonium ferrous sulfate (Mohr's salt) were from Fluka.

The buffer HEPES and sodium phosphate were from Fluka.

All the molecular graphics pictures was generated with PyMOL software (DeLano Scientific ltd)[38] .Data analysis was made under Microcal Origin software.

Chapter 2: De novo design of DF-Click

De novo design of DF-Click

Design of the linker was accomplished by means of Accelrys Discovery Studio software, starting from manually created coordinates of the linker and NMR derived coordinates of DF3 (PDB-ID: 2kik). Basically, once determined distances and compatible positions for propargyl glycine insertion, a simple routine for superimposition was adopted. Tethers were introduced between main chain atoms of Pra and the residue under evaluation: acyl atoms of 6-aha on the one side of the linker and acetyl ones on the other side. Free torsions around all the aliphatic bonds of the linker allowed the superimposition task for all the three positions evaluated. The obtained conformers obtained were evaluated by a simple Dreiding potential and ranked in terms of improper dihedrals.

Design of DF-Click1 was accomplished with Accelrys Discovery Studio software. Starting from the DF1-L13G (PDB-ID: 1lt1) crystal structure coordinates the first step was to introduce mutations as in DF3. Introduction of the loop residues was accomplished in one step and several minimization cycles were performed, while keeping helical residues restrained in this phase of the design. Glycine residues were then inserted in position 9 and 13, for both subunits without restraints. All the subsequent mutations were introduced after minimization cycles. A final adopted basis Newton-Raphson minimization cycle was accomplished for 2000 steps, with two distance restraints between zinc

and histidine Nδ coordinating atom. DF-Click1 molecule was typed with all-atom Discovery Studio implementation of CHARMM forcefield, a simplified Generalized Born method was adopted for solvation energies and forces, long-range contribution were accounted with a 1.4 nm spherical cutoff with an exponential switching starting from 1.2 nm. Swapped DF-Click1 was modeled starting from minimized DF-Click1 coordinates in Pymol, by manual searching the best rotamers for Phe and Thr.

Solid phase peptide synthesis: overview

Since R. B. Merrifield (Nobel prize in 1984) reported on the solid phase chemical synthesis of peptides and small proteins, solid-phase peptide synthesis (SPPS) has become a widely utilized technique for the production of peptides and small proteins[25]. The basic idea of the solid-phase approach involves the covalent attachment (anchoring) of the growing peptide chain to an insoluble polymeric support (resin); in this way unreacted soluble reagents can be removed simply by filtration and washing, without losses of the growing peptide chain. Subsequently, the insoluble peptide-resin is elongated by addition of α-amino- and side-chain-protected amino acid residues in a series of addition reactions, which must be performed with high yields. Excess soluble reagents are used to drive sequential coupling and deprotection reactions to completion. The repetitive nature of the coupling and deprotection steps allowed the automation of the main steps of the SPPS methodology.

Chapter 2: De novo design of DF-Click

The acid-labile Boc group, or base-labile Fmoc-group, is used for the N-α-protection of amino acids. The general scheme of SPPS using Fmoc chemistry is illustrated in Figure 2.20. Fmoc- protected amino acids are most popular in SPPS, since a mild basic treatment, using piperidine, can be used for the Fmoc-group removal at every stage. The final cleavage and deprotection of the peptidyl resin is accomplished using TFA. An N-α-derivatized amino acid is attached to an insoluble (solid) support, via a flexible linker. The N-α-blocking group is then removed (deprotected) and the amino acid-linker-support is thoroughly washed with solvent. The next amino acid (which is also N-α-protected) is then coupled to the previous amino acid.

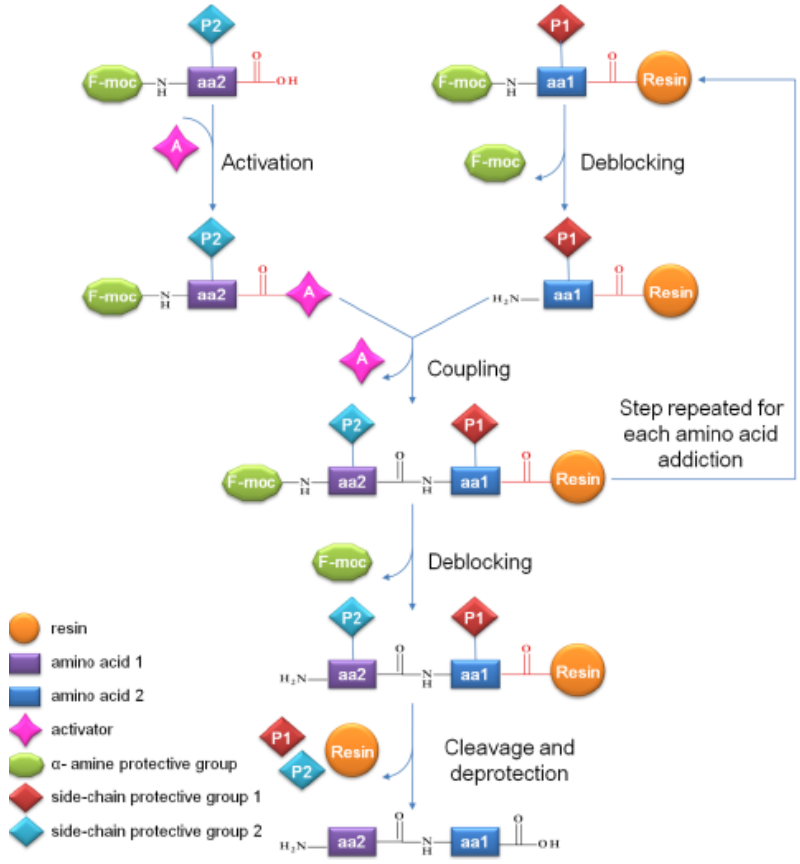


Figure 2.20. General scheme of solid-phase peptide synthesis.

The deprotection/coupling cycle is repeated until the desired sequence is completed. The peptide-linker-support is cleaved to obtain the peptide as a free acid or amide, depending on the chemical nature of the linker. The cleavage reagent also removes the amino acid side chain protecting groups, which are unstable to the deprotection reagent. All the synthetic steps may be carried out either in batch process, where the support is filtered between each step, or in a continuous flow process, where the support is always solvated during

reagent exchange, and the release of the protecting groups is spectrophotometrically monitored.

Solid phase synthesis of peptides: Chain K and Chain Z

The 48 amino acid peptides, Z- and K- DF-Click1 were synthesized by automatic solid-phase synthesis using an ABI 433A peptide synthesizer (Applied Biosystem, Foster City, CA, USA) with Fmoc-protocols on a 0.1 mmol scale. The N- and C-termini were acetylated and amidated, respectively, except for Z chain, which has been N-capped with 6-azidohexanoic acid, as described below. The resin used was acid labile H-PAL ChemMatrix (Figure 2.21), with a substitution level 0.31 mmol/g. This resin is particularly suitable for the synthesis of large or hydrophobic peptides and small proteins.

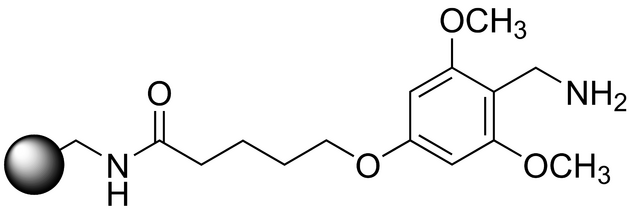


Figure 2.21. H-PAL amide resin for SPPS

The following protected amino acids were adopted:

- **Chain K**: Fmoc-Asp(OtBu)-OH; Fmoc-Glu(OtBu)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Leu-OH; Fmoc-Ile-OH, Fmoc-Gly-OH; Fmoc-Ala-OH; Fmoc-Val-OH; Fmoc-Pro-OH; Fmoc-Phe-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(tBu)-OH;

Fmoc-Thr(tBu)-OH; Fmoc-His(Trt)-OH Fmoc-Lys(Boc)-OH; Fmoc-Arg (Pbf)-OH ; Fmoc- propargylglycine –OH**;**

- **Chain Z**: Fmoc-Asp(OtBu)-OH; Fmoc-Glu(OtBu)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Leu-OH; Fmoc-Ile-OH, Fmoc-Gly-OH; Fmoc-Ala-OH; Fmoc-Val-OH; Fmoc-Pro-OH; Fmoc-Phe-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(tBu)-OH; Fmoc-Thr(tBu)-OH; Fmoc-His(Trt)-OH Fmoc-Lys(Boc)-OH; Fmoc-Arg (Pbf)-OH**.**

The synthetic procedure can be summarized as follows:

1. **Deprotection**: Fmoc group was removed with a 20% piperidine solution. After Fmoc deprotection, a NMP washing module was applied to the reaction.

2. **Activation**: the carboxyl group of each Fmoc-amino acid was activated by direct addition in cartridge of 1 mmol HCTU. HATU was adopted for difficult couplings.

3. **Coupling**: the pre-activated Fmoc-amino acid reacted with the amino-terminal group of the growing peptide chain on the resin in NMP.

4. **Capping**: this reaction was performed after each coupling step, using Ac_2O/HOBt/DIEA solution in NMP. Capping cycle was introduced to prevent deletion byproducts.

When the synthesis was complete, the resin was washed several times with DCM, NMP, isopropanol and methanol, and finally dried.

Coupling of Chain Z with 6-Azido hexanoic acid for Click Chemistry reaction

After the final coupling step, chain Z was functionalized with 6-azidohexanoic acid on the N-terminal amino group by using peptide automatic synthesizer.

The coupling reaction was performed by dissolving 1 mmol of HATU with 1 mmol of 6-azidohexanoic acid in NMP, directly in cartridge. This solution was then transferred to the reaction vessel, and the reaction mixture was incubated at room temperature for 40 minutes.

Peptides cleavage from the resin and side chain deprotection

The peptide cleavage from the resin and side chain deprotection was carried out using a solution of 95% TFA, 2.5% TIS, 2.5% H_2O (v/v/v 20 mL). The reaction was carried out for one hour at 0 ° C and one hour at room temperature, under a slight magnetic stirring.

The resin was filtered and the solution was collected under vacuum in a glass flask, than the resin was washed again with undiluted TFA, which was removed by rotary evaporation.

The crude peptide was precipitated immediately, by adding 5 volumes of cold diethyl ether to the concentrated reaction mixture. The mixture was centrifuged, the supernatant was removed and the precipitate was washed twice with three volumes of cold diethyl ether.

The peptide was dried to remove diethyl ether, re-dissolved in water and acetonitrile and lyophilized.

Analysis and purification of chain Z

The identity of crude peptide Z was analyzed by analytical RP-HPLC–ESI-MS. The LC-MS was performed with a Shimadzu LC-10ADvp equipped with an SPDM10Avp diode-array detector. ESI-MS spectra were recorded on a Shimadzu LC-MS-2010EV system with ESI interface and Shimadzu LC-MS solution Workstation software for the data processing.

A Q-array-octapole-quadrupole mass analyzer was used as the detector.

All analyses were performed with a Vydac C18 column (4.6 mm x 150 mm; 5μm), eluted with an H_2O 0.1 % trifluoroacetic acid, TFA, (eluent A) and CH_3CN 0.1 % TFA (eluent B) linear gradient, from 30 to 62% (solvent B), over 32 minutes, at 0.5 ml min^{-1} flow rate. The optimized MS parameters were selected as followed: CDL (curved desolvation line) temperature 250 °C; the block temperature 250 °C; the probe temperature 250 °C; detector gain 1.6kV; probe voltage +4.5kV; CDL voltage -15V. Nitrogen served as nebulizer gas (flow rate: 1.5 $L \cdot min^{-1}$).

The purification of fully deprotected peptide was accomplished by preparative RP-HPLC with Shimadzu LC-8A connected to a SPD-10A Shimadzu UV-Vis spectrophotometric detector. Chromatographic column Vydac C18 (22 mm x 250 cm; 10 μm) was used at a flow rate of 21 $mL \cdot min^{-1}$, with a linear gradient, appropriately modified and

scaled. The identity of the peak was verified by LC-MS. The pooled fractions, containing the desired products, were lyophilized. The Peptides homogeneity was assessed by analytical HPLC and by ESI mass spectrometry (Figure 2.22).

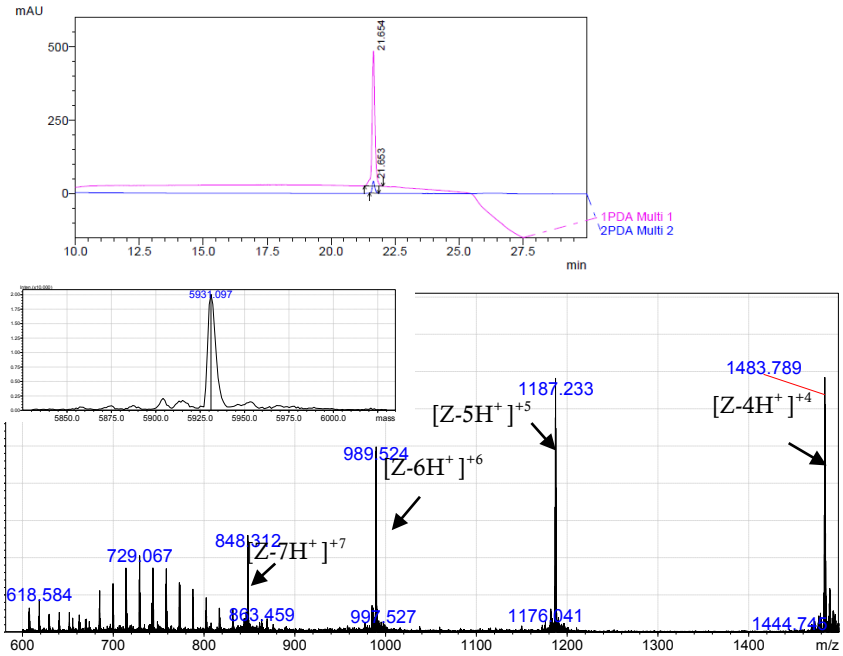


Figure 2.22. RP-HPLC chromatogram of pure peptide Z; ESI-TOF spectrum relative to the peak at R_t= 21.654 min; deconvolution for previous spectrum (theoretical mass: 5931.71 Da; observed mass: 5931.10 ± 0.6 Da) is in the inlet.

Analysis and purification of chain K

The fully deprotected peptide was analyzed by RP-HPLC-ESI-MS. The mobile phase, at a flow rate of 0.5 mL·min^{-1}, was made up of H_2O 0.1% TFA (solvent A) and CH_3CN/0.1% TFA (solvent B) linear

gradient, from 10 to 62% (solvent B), over 32 min. The Vydac C18 column (4.6 mm x 150 mm; 5µm) was used for analyses.

Peptide K was further purified to homogeneity by preparative RP-HPLC to a single, symmetric peak by Vydac C18 (22 mm x 250 cm; 10 µm) at a flow rate of 21 mL min^{-1}, with a gradient of acetonitrile in 0.1% aqueous TFA, from 10% to 62%, over 27 minutes. The collected fractions were analyzed by RP-HPLC-ESI-MS, mixed, lyophilized, and stored at -20°C (Figure 2.23).

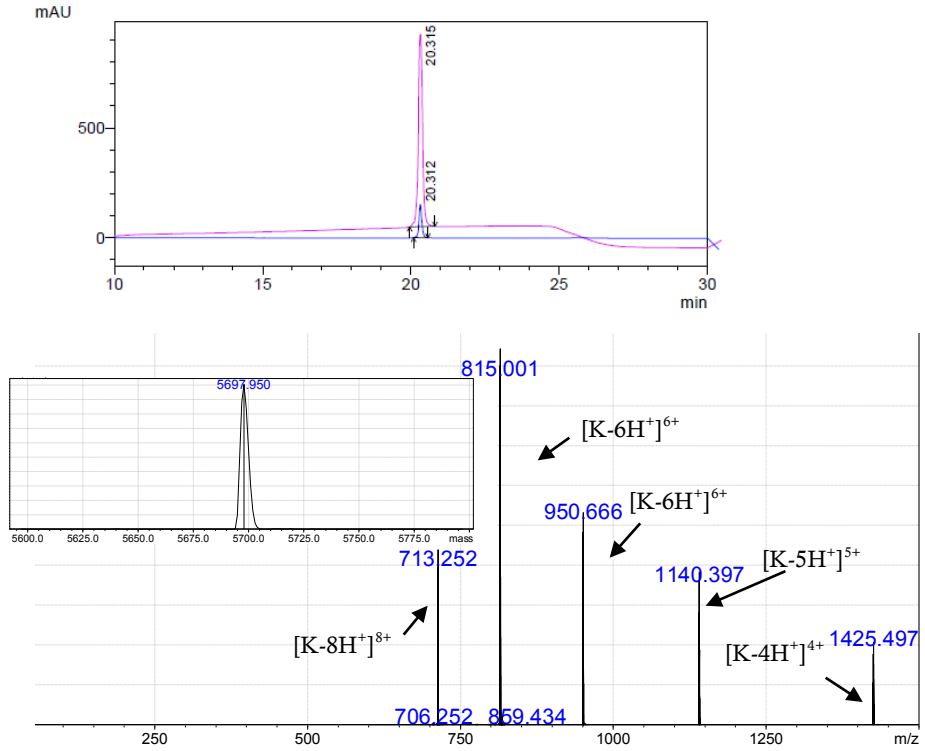


Figure 2.23. RP-HPLC chromatogram of pure peptide K ; ESI-TOF spectrum relative to the peak at R_t= 20.315 min; deconvolution for previous spectrum (theoretical mass: 5698.34 Da; observed mass: 5697.95 ± 0.4 Da) is in the inlet.

Chapter 2: De novo design of DF-Click

Click Chemistry reaction: DF-Click1

Chain K and Chain Z peptides were covalently conjugated in solution by Click chemistry. The heterodimer DF-Click1 formation was obtained in the following steps:

- Chain Z (14 mg; $2.04 \cdot 10^{-3}$ mmol; MW = 6835,69 mg/mmol) was mixed with an excess (1.5 eq) of Chain K (19,4 mg; $3,07 \cdot 10^{-3}$ mmol; MW = 6302,34 mg /mmol) in 7.5 mL HEPES buffered solution (200 mM) at pH= 8.0.
- The solution was filtered and placed into a flask, under inert atmosphere, and the reaction was carried out with magnetic stirring.
- The catalyst solution was prepared by dissolving 2.5 eq of $CuSO_4 \cdot 5H_2O$ (28.27 mg; 0,11 mmol; MW = 249.68 mg/mmol) and 5.0 eq of TCEP (64.03 mg; 0,22 mmol; MW = 286.65 mg/mmol) in 500 μL of HEPES buffer solution at pH=8.
- Upon fading of this solution (indicating the Cu(II) reduction), 2.5 eq of catalyst solution were added (25 μL) to the reaction mixture. The solution immediately turned to brown/ocher.

The reaction was followed by analytical RP-HPLC and product formation was ascertained by ESI-MS (Figure 2.24). The analysis of the chromatogram and the mass spectra showed the successful formation of DF-Click1 in 3.5 hours, with only a slight increase in calculated yield upon a further addition of the catalyst. Product purification was performed by preparative RP-HPLC on a 2.2 cm

Vydac C18 column, at a flow rate of 21 mL min^{-1}, with a gradient of acetonitrile in 0.1% aqueous TFA, from 30% to 62% over 32 min.

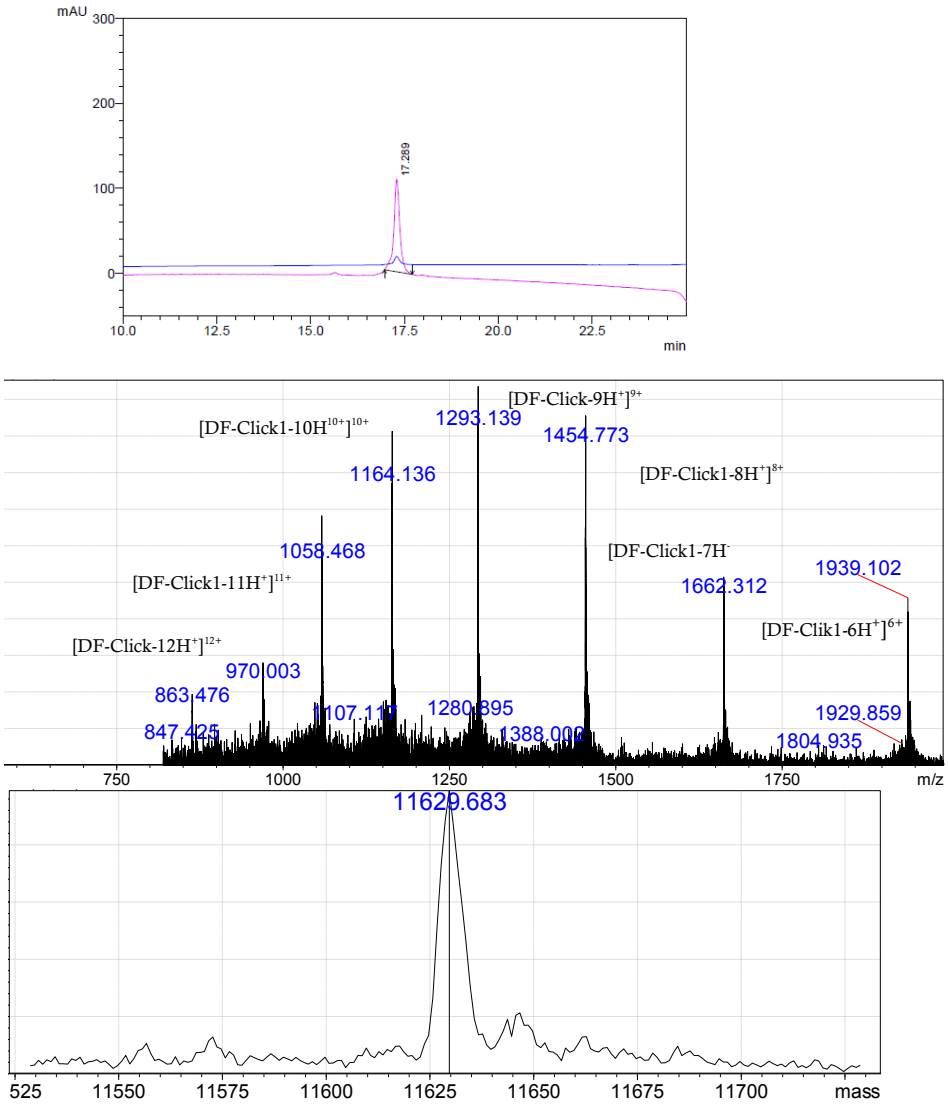


Figure 2.24. RP-HPLC chromatogram of pure DF-Click1 ; ESI-TOF spectrum relative to the peak at R_t= 17.289 min; deconvolution for previous spectrum (theoretical mass: 11630.05 Da; observed mass: 11629.68 ± 0.4 Da).

The pooled fractions, containing the desired product, were lyophilized. The final yield of pure product was of 63%, based on theoretical 1:1 monomer ligation.

NMR Experiments and diffusion ordered spectroscopy DOSY-NMR

NMR spectroscopy was used to analyze the Apo-DF-Click1 and the Zn-DF-Click1. All solutions for NMR spectra were prepared with 10 % of D_2O. The apo form was analyzed at concentration of 500 μM in H_2O. Zn-DF-Click1 complex was obtained by adding 10 eq of a $ZnCl_2$ aqueous solution to a 130 μM of apo form in phosphate buffer pH 6.5. The solution was allowed to equilibrate for 15 minutes and, after the concentration, was increased to 400 μM using Vivaspin filter (5,000 molecular weight cutoff) (Vivascience, Edgewood, NY, USA). All NMR spectra were acquired at 25 °C, on a Bruker Avance 600 spectrometer equipped with a triple resonance cryo-probe. Suppression of the water signal was accomplished by excitation sculpting sequence[39].

2D-DOSY spectrum of the Zn(II) complex (0.4 mM) was obtained in a 50mM NaP buffer pH 7.0. Exponential fitting of the signal decay was made with the Bruker proprietary software. 128 scans were recorded for each FID, with total acquisition time 4h for the bdLED-DOSY experiment. The parameters for DOSY diffusion time are Δ=100ms δ=2.5ms τ=0.5ms.

Hydrodynamic calculations were done, by using glycerol as an internal standard[40]. The theoretical value for DF-Click1 was calculated by the shell model with hydration layer 3.2 Å of HYDRONMR, with parameters corrected according to experimental conditions, starting from the minimized model coordinates[29].

Metal binding experiments

UV-Vis analysis

UV–Vis spectra were recorded with a Cary Varian 50 Spectrophotometer, equipped with a thermostated cell compartment (Varian, Palo Alto, CA, USA) using quartz cuvettes with 1 cm or 0.1 cm path lengths. Wavelength scans were performed at 25 °C from 200 to 800 nm, with a 60 $nm \cdot min^{-1}$ scan speed. Extinction coefficients are expressed per dimetal site, unless otherwise noted.

Preparation of Co(II)-DF-Click1 metal complexes

The Co(II)-DF-Click1 complex was prepared from a stock solution of apo DF-Click1 in HEPES buffer 50 mM, 100 mM NaCl, pH 7.0 by adding 2 μL of a solution of $CoCl_2$ ($5.6 * 10^{-2}$ M in H_2O 0.005% TFA).

The initial DF-Click1 concentration (73.6 μM) was spectrophotometrically determined, by using ε_{280}=15.470 cm^{-1} M^{-1}. Metal stock solutions were standardized by ICP-MS analytical methods. Both protein and metal stock solution were freshly prepared.

Preparation of Fe(III)-DF-Click1 metal complexes

Fe(III)-DF-Click1 complex was obtained upon addition of stoichiometric amount of Mohr's Salt (in 0.05% H_2SO_4 water solution)

to a solution of the apo-dimer (148 μM) in HEPES buffer 50 mM, 100 mM NaCl pH 7.0 , under Ar atmosphere. The mixture was exposed to air for 1h at room temperature, under magnetic stirring. A small quantity of precipitates, presumably iron oxides and aggregated protein, appeared. Then, the solution was centrifuged on a Vivaspin filter (5,000 molecular weight cutoff), after several steps of washing with appropriate buffer solution.

Circular dichroism analysis

Circular dichroism (CD) measurements were performed using a J-815 spectropolarimeter equipped with a thermostated cell holder (JASCO, Easton, MD, USA). CD spectra were collected at 25°C, from 260 to 190 nm at 0.2 nm intervals with a 20 nm min^{-1} scan speed, at 1 nm band width and at 16 s response. Cells of 0.5 cm path length were used in all measurements.

Mean residue ellipticities [θ] were calculated using the equation $[\theta]=\theta_{obs}/(10 \cdot l \cdot C \cdot n)$, in which θ_{obs} is the ellipticity measured in millidegrees, l is the path length of the cell in centimeters, C is the concentration in moles per liter, and n is the number of residues in the protein.

Metal ion titrations were carried out by adding small aliquots of freshly prepared aqueous stock solutions of MCl_2 (M is Zn, Co, Mn, Ni and Cu) to the protein solution (concentration of 5.20 μM per dimer, in 10 mM phosphate buffer at pH 7), under magnetic stirring. Samples were allowed to incubate for 5 min before measurement.

The iron ion titration was carried out under magnetic stirring and inert Ar atmosphere.

The protein solution was prepared at a concentration of 5.20 μM per dimer, by dilution of a stock solution (51.7 μM in H_20 0.05 % TFA) with 10 mM phosphate buffer at pH 7. This solution was lyophilized and then re-dissolved in water, previously degassed with an Ar stream, for one hour. The titration was carried out by adding small amounts of Mohr' Salt, previously degassed.

The equation for the CD data fitting was obtained by considering the following equilibrium for metals binding:

$$P + nM \rightleftarrows M_nP$$

The relative dissociation constant was:

$$K_d = \frac{[P][M]^n}{[M_nP]}$$

where [P] and [M] were the concentration of free protein and free metal, respectively. These can be expressed as:

$$P_T = [P] + [M_nP] \rightarrow [\boldsymbol{P}] = \boldsymbol{P_T} - [\boldsymbol{M_nP}]$$

$$M_T = [M] + n[M_nP] \rightarrow [\boldsymbol{M}] = \boldsymbol{M_T} - \boldsymbol{n}[\boldsymbol{M_nP}]$$

where [P]T, [M]T , [MnP] ware the concentration of total protein, total metal and complex, respectively.

$$\alpha = \alpha_P[P] + \alpha_{M_nP}[M_nP]$$

$$\alpha_0 = \alpha_P P_T \Rightarrow \alpha_P = \alpha_0/P_T$$

$$\alpha_\infty = \alpha_{M_nP} P_T \Rightarrow \alpha_{M_nP} = \alpha_\infty/P_T$$

in which α, α_0 and α represent the ellipticity at the actual metal concentration, at zero metal concentration and metal saturation.

Than, substitution in α makes it dependent on the concentration of the complex:

$$\alpha = \left(\frac{\alpha_0}{P_T}\right)(P_T - [M_nP]) + \left(\frac{\alpha_\infty}{P_T}\right)[M_nP]$$

$$\alpha P_T = \alpha_0 P_T + (\alpha_\infty - \alpha_0)[M_nP]$$

$$P_T \frac{(\alpha-\alpha_0)}{(\alpha_\infty-\alpha_0)} = [M_nP] \qquad (1)$$

K_d can be arranged as follows:

$$K_d = \frac{(P_T - [M_nP])(M_T - n[M_nP])^n}{[M_nP]}$$

$$K_d[M_nP] = (P_T - [M_nP])(M_T - n[M_nP])^n$$

$$\frac{K_d[M_nP]}{(P_T - [M_nP])} = (M_T - n[M_nP])^n$$

$$\left(\frac{K_d[M_nP]}{(P_T - [M_nP])}\right)^{\frac{1}{n}} = M_T - n[M_nP]$$

$$n[M_nP] + \left(\frac{K_d[M_nP]}{(P_T - [M_nP])}\right)^{\frac{1}{n}} = M_T$$

$$n[M_nP] + \left(\frac{K_d}{\left(\frac{P_T}{[M_nP]} - 1\right)}\right)^{\frac{1}{n}} = M_T \quad (2)$$

Replacing eq. (1) in eq. (2):

$$M_T = nP_T\frac{(\alpha - \alpha_0)}{(\alpha_\infty - \alpha_0)} + \frac{K_d^{\frac{1}{n}}}{\left(\frac{1}{\frac{(\alpha - \alpha_0)}{(\alpha_\infty - \alpha_0)}} - 1\right)^{\frac{1}{n}}}$$

Finally this equation was implemented and used in the MicroCal Origin software, which gave the expected protein/metal ratio of 1:2.

References

(1) Marsh, E. N. G.; DeGrado, W. F. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5150.

(2) Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; Degrado, W. F. *Nat. Chem.* **2012**, *4*, 900.

(3) Summa, C. M.; Rosenblatt, M. M.; Hong, J.-K.; Lear, J. D.; DeGrado, W. F. *J. Mol. Biol.* **2002**, *321*, 923.

(4) Kaplan, J.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11566.

(5) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.

(6) Huisgen, R. *Proc. Chem. Soc.* **1961**, 357.

(7) Meldal, M.; Tornøe, C. W. *Chem Rev* **2008**, *108*, 2952.

(8) Hein, C. D.; Liu, X.-M.; Wang, D. *Pharm. Res.* **2008**, *25*, 2216.

(9) Brik, A.; Wu, C.-Y.; Wong, C.-H. *Org. Biomol. Chem.* **2006**, *4*, 1446.

(10) Breinbauer, R.; Köhn, M. *ChemBioChem* **2003**, *4*, 1147.

(11) Ciferri, A.; Perico, A. *Ionic Interactions in Natural and Synthetic Macromolecules*; John Wiley & Sons, 2012.

(12) Samish, I.; MacDermaid, C. M.; Perez-Aguilar, J. M.; Saven, J. G. *Annu. Rev. Phys. Chem.* **2011**, *62*, 129.

(13) Gonzalez, L.; Woolfson, D. N.; Alber, T. *Nat. Struct. Mol. Biol.* **1996**, *3*, 1011.

(14) Punna, S.; Kuzelka, J.; Wang, Q.; Finn, M. G. *Angew. Chem. Int. Ed.* **2005**, *44*, 2215.

(15) Humenik, M.; Huang, Y.; Wang, Y.; Sprinzl, M. *ChemBioChem* **2007**, *8*, 1103.

(16) Waengler, C.; Schaefer, M.; Schirrmacher, R.; Bartenstein, P.; Waengler, B. *Bioorg. Med. Chem.* **2011**, *19*, 3864.

(17) Thonon, D.; Kech, C.; Paris, J.; Lemaire, C.; Luxen, A. *Bioconjug. Chem.* **2009**, *20*, 817.
(18) Faiella, M.; Andreozzi, C.; de Rosales, R. T. M.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A. *Nat Chem Biol* **2009**, *5*, 882.
(19) Torres Martin de Rosales, R.; Faiella, M.; Farquhar, E.; Que, L.; Andreozzi, C.; Pavone, V.; Maglio, O.; Nastri, F.; Lombardi, A. *JBIC J. Biol. Inorg. Chem.* **2010**, *15*, 717.
(20) Maglio, O.; Nastri, F.; Martin de Rosales, R. T.; Faiella, M.; Pavone, V.; DeGrado, W. F.; Lombardi, A. *Comptes Rendus Chim.* **2007**, *10*, 703.
(21) Murray, L. J.; Lippard, S. J. *Acc Chem Res* **2007**, *40*, 466.
(22) Elsen, N. L.; Bailey, L. J.; Hauser, A. D.; Fox, B. G. *Biochemistry (Mosc.)* **2009**, *48*, 3838.
(23) Sazinsky, M. H.; Merkx, M.; Cadieux, E.; Tang, S.; Lippard, S. J. *Biochemistry (Mosc.)* **2004**, *43*, 16263.
(24) Mitchell, K. H.; Studts, J. M.; Fox, B. G. *Biochemistry (Mosc.)* **2002**, *41*, 3176.
(25) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149.
(26) Atherton, E.; Sheppard, R. C. *Solid phase peptide synthesis: a practical approach*; IRL Press, 1989.
(27) *Protein NMR spectroscopy: principles and practice*; Cavanagh, J., Ed.; 2nd ed.; Academic Press: Amsterdam ; Boston, 2007.
(28) Jones, J. A.; Wilkins, D. K.; Smith, L. J.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 199.
(29) García de la Torre, J.; Huertas, M. L.; Carrasco, B. *J. Magn. Reson.* **2000**, *147*, 138.
(30) Torres, O.; Yüksel, D.; Bernardina, M.; Kumar, K.; Bong, D. *Chembiochem Eur. J. Chem. Biol.* **2008**, *9*, 1701.
(31) Calhoun, J. R.; Kono, H.; Lahr, S.; Wang, W.; DeGrado, W. F.; Saven, J. G. *J. Mol. Biol.* **2003**, *334*, 1101.
(32) Nastri, F.; Bruni, R.; Maglio, O.; Lombardi, A. In *Coordination Chemistry in Protein Cages*; Ueno, T.; Watanabe, Y., Eds.; John Wiley & Sons, Inc., 2013; pp. 43–85.
(33) Dassama, L. M. K.; Krebs, C.; Bollinger, J. M., Jr; Rosenzweig, A. C.; Boal, A. K. *Biochemistry (Mosc.)* **2013**, *52*, 6424.

(34) Lippard, S. J.; Berg, J. M. *Principles of bioinorganic chemistry*; University Science Books: Mill Valley, Calif., 1994.
(35) Bertini, I.; Luchinat, C. *Adv. Inorg. Biochem.* **1984**, *6*, 71.
(36) Strand, K. R.; Karlsen, S.; Andersson, K. K. *J. Biol. Chem.* **2002**, *277*, 34229.
(37) Solomon, E. I.; Brunold, T. C.; Davis, M. I.; Kemsley, J. N.; Lee, S.-K.; Lehnert, N.; Neese, F.; Skulan, A. J.; Yang, Y.-S.; Zhou, J. *Chem. Rev.* **2000**, *100*, 235.
(38) Schrödinger, L. L. C. The PyMOL Molecular Graphics System, Version 1.3r1, 2010.
(39) Hwang, T. L.; Shaka, A. J. *J. Magn. Reson. A* **1995**, *112*, 275.
(40) SCHULTZ, S. G.; SOLOMON, A. K. *J. Gen. Physiol.* **1961**, *44*, 1189.

Chapter 3: *De novo* design of QF

homotetramer, four metal ions, protein crystal design, X-ray

3 *De novo* design of QF

3.1 Results and discussion

The Design of a single peptide chain, which is able to self assemble in a four-helix bundle and efficiently binds up to four metal ions is an intriguing challenge. Some successful attempts already appeared in the literature. Ogawa and co-workers, reported the Cu(I) binding properties of a designed peptide, called C16C19-GGY[1,2]. A CxxC motif is responsible of the formation of a multinuclear Cu(I) complex as confirmed by EXAFS analysis. This compound is stable in reducing conditions and a 1:1 stoichiometry was confirmed by mass spectrometry. Its spectroscopic features are very similar to cysteine rich metallothionein, which are also able to stabilize multinuclear copper complexes.

Nanda and co-workers reported the *de novo* design of a $[Fe_4S_4]$ cluster in a non-biological four-helix bundle environment[3]. Making use of C_2 symmetry as in the DF design, the stepwise design process gave a monomeric four-helix bundle able to bind the Fe-S cluster in its hydrophobic core. CCIS1 was able to fold in the intended secondary

structure, as assessed by CD and to bind [Fe_4S_4], as assessed by UV and EPR spectroscopies. However, some unexpected oligomerization, and the inability to cycle between reduced and oxidized forms, limited the study of this compound.

Nature adopts metal clusters in several processes: electron-transfer[4], molecular hydrogen uptake and production[5], nitrogen fixation[6] are only few noteworthy examples. Manganese cluster of Photosystem II (PSII) is unique in its ability to catalyse four electrons water oxidation in its oxygen-evolving complex (OEC).

$2H_2O = O_2 + 4e^- + 4H^+ \quad E^0 = -1.23V$ (vs NHE)

Water oxidation standard potential is -1.23 V, one of the lowest in nature, and water splitting is accomplished in a mechanism that involves four intermediate states, in which manganese ions cycle between III and IV oxidation states[5,7].

High-resolution X-ray structural data were crucial to shed light on the mechanism of this complex biological machinery. Previously cited protein design work had severe limitations in the structural analysis of the metal complex, because of the poor structural stability of the apo-form. For this reason, in this work the design approach was crystal-oriented.

Lanci et al. reported for the first time the *de novo* design of a protein crystal[8]. A P6 crystal structure was obtained as intended in the design for a trimeric coiled-coil peptide, by carefully evaluating the interactions among several trimeric units. Prediction of the preferred

topologies was accomplished through single point energy calculations; iteration on two degrees of freedom (inter-bundle distance and mutual angle) produced a final matrix of the sequence-structure energy landscape that allowed identifying the best sequence.

In this work, the *de novo* design of a new class of synthetic metalloproteins is presented, inspired to the tetranuclear core of the OEC of the PSII. The QF (*Quattro Ferro*) project involves the design of the core unit, which was adopted for the design of four different analogues for solution and crystal characterization. A combination of circular dichroism, UV-Vis and NMR spectroscopies were adopted to characterize the synthesized analogues, and the crystal structure of the apo-form of one of them is reported.

3.1.1 Design

Design of a tetranuclear metal cofactor

QF core *de novo* design was realized via a stepwise process. The choice of the four-helix bundle as scaffold derives from our previous work in the dinuclear metal binding *de novo* designed proteins, called DF. As already pointed out in the introduction of this thesis, four-helix bundle is the perfect candidate for our design process; it is highly stable, finely tunable and robust toward several mutations, making it 'designable'[9–11]. In this work, a more stringent D_2 symmetry was adopted, in order to simplify calculations during the design process and the synthesis of several analogues.

Chapter 3: De novo design of QF

The first step of the design process consisted in the generation of a backbone, which would be able to host up to four metal ions. For this purpose the DF1 backbone appeared a perfect candidate[12–14]. As we are interested in the binding of more than two metals, more ligands are needed, with respect to DF. For this reason, we generated a tetramer starting from the sole helix 2 and 2' of DF1, both bearing the ExxH binding motif. In order to fulfil the symmetry requirement, the 2.5 Å Zn complex crystallographic structure was adopted (PDB ID: 1ec5). Chain A in this crystal form generates a perfect dimer by means of a crystallographic C_2 operation (Figure 3.1).

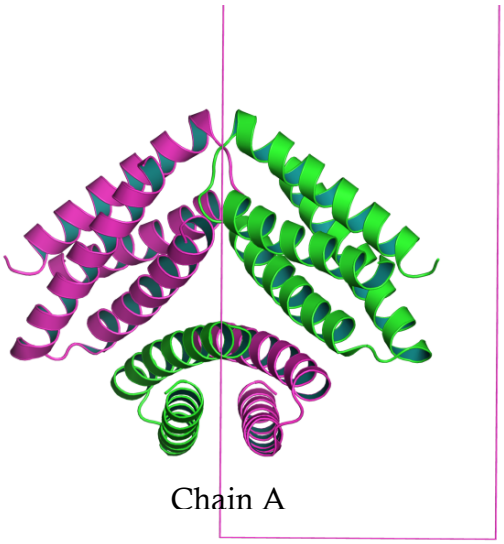


Figure 3.1. Asymmetrical unit (green) and its crystallographic symmetry mate (magenta) of Zn(II)-DF1 crystal structure (PDB ID: 1ec5). Magenta axes represent the unit cell.

Helices 2 and 2' were extrapolated (from residues 26 to 47), and a second C_2 symmetry was applied to generate the anti-parallel four-helix

bundle. All DF1 residues were mutated to Ala, except for the key residues, involved in the first and second coordination shells: Glu 36, His 39, Asp 35 and Lys 38.

Even though, a good set of coordinates for the backbone was obtained, the aim is to completely design a protein from a scratch. For this reason, Coiled Coil Crick parameterization server (CCCP)[9] was adopted to fit the DF derived backbone. Fitting was excellent (<0.2 Å), and the parameters obtained (Table 3-1) were adopted to generate a three heptad repeat (27 residues) four-helix bundle.

Table 3-1. Crick parameters obtained from the fitting of a tetramer composed by the helix 2 of DF.

R0 (Å)	7.596
R1 (Å)	2.231
ω_0 (°/res)	-3.130
ω_1 (°/res)	102.483
α (°)	-15.827
ϕ_1 (°)	-6.002
$\Delta\phi_0$ of D_n symmetry (°)	-96.969
starting heptad position	f
pitch (Å)	168.361
rise per residue (Å)	1.521

Backbone from the key residues responsible of the metal binding was kept unaltered, and constrained regularization allowed binding

this portion to the CCCP-parameterized coiled coil. At this point, residues in *a*, *d*, positions responsible for the internal hydrophobic core, were designed *via* computational routines for side chain assignments[15,16]. External positions were intentionally not inserted in the packing, since they are involved in the crystal interactions between different tetramers. Tightly packed Leu and Ile in *a* and *d* positions occupy the interior of the protein, except for the ExxH motif.

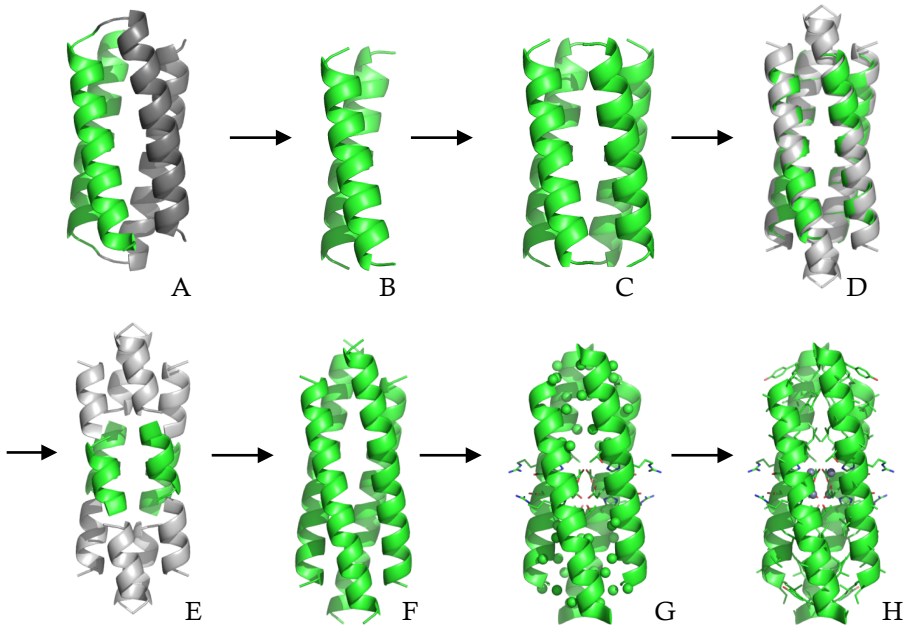


Figure 3.2. Pictorial description of QF design process: Zn(II)-DF1 crystallographic dimer (PDB access code 1EC5) was chosen as the starting template (A); helices 2 and 2' bringing the EXXH binding motif were extracted for tetramer formation (B,C); the obtained tetramer was fitted to a Coiled-Coil Crick parameterization software 2 in order to assure an ideal four-helix bundle (D); metal binding heptad was preserved and backbone relaxed in order to host the metals (E,F); first and second sphere coordination residues were included and the packing of the side chains was accomplished, by means of Rosetta fixed backbone design software.

A final step of molecular mechanics minimization was performed to relax the whole structure.

Molecular mechanics analysis allowed us to hypothesize a geometry of the ligands, where each metal ion is bound to one histidine residue and three glutamates, each one involved in a μ^3-η^1:η^2 carboxylate bridging, with a distorted tetrahedral geometry (Figure 3.3).

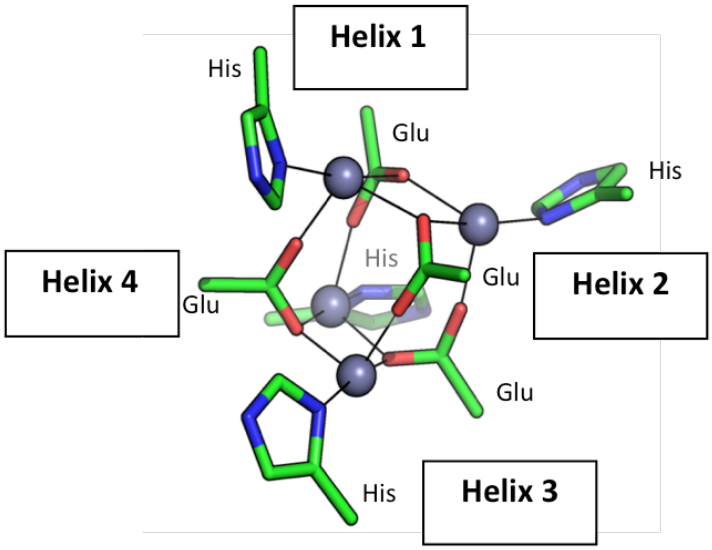


Figure 3.3. Tridimensional model of the tetranuclear metal cofactor in QF.

De-novo design of a metal binding protein crystal

In order to build a soft-matter nanodevice, we started the design of the QF in a protein crystal context. The recent protocol for protein crystal design[8] may lead to successful protein crystallization with a desired topology in these following steps: (i) determination of the constituent protein, (ii) generation of a set of accessible crystalline

arrays of the backbone, (iii) identification of candidate crystalline structures within this set, (iv) design of sequences for these backbones, (v) experimental test of best candidates.

In the topology chosen for this design process, a polar crystalline order was targeted in order to give nonlinear optical properties to the resulting crystal. As reported by Lanci et al. [8], the target topology would include large water channels, in which future positioning of cofactors and guest molecules may arise in the construction of light harvesting devices. For this reason, a layered, porous $P4_1$ symmetry, in which a tetramer is the asymmetrical unit, would present some advantages: the high symmetry facilitates the design process and the acquisition of X-ray data, the tubular structure alternates voids in the crystal where external cofactor may soak, the already cited polarity.

In order to limit the number of degrees of freedom in the generation of possible crystalline configurations for a $P4_1$ symmetry: (i) The C_2 axis parallel to the bundle growth coincides to the principal screw axis in the crystal, (ii) the angle between two bundles in the same plane is fixed to give only one designable interface. The only two degrees of freedom left are **a**=**b** and **c** cell edges lengths. The tubular orientation consists of pseudo-contiguous super helix and hydrophobic core. To drive the desired self-assembly, each helix must span a multiple of $\pi/2$ about the principal screw axis, and so only some peptide chain lengths are allowed:

$$\omega_0\, t = n\frac{\pi}{2} \qquad\qquad n = 1,2,\dots,N$$

where ω_0 is the superhelical frequency, t is the number of residues. Given a frequency of 3.130 °/res, as previously fitted with CCCP, the number of residues should be approximately 29 for n=1. Considering N-termini and C-termini capping acetyl and amide, a 27 residue peptide would allow for the necessary rearrangement at the edges, while keeping intact the intended tubular structure.

Generation of the candidate crystalline arrays of the backbone, by iteration of the degrees of freedom, give rise to a grid-search, where **a** edge goes from 21 to 28 Å in 0.2 Å increments, and **c** edge goes from 154 to 170 Å in 2 Å increments (324 structures, 36x9 matrix) (Figure 3.4).

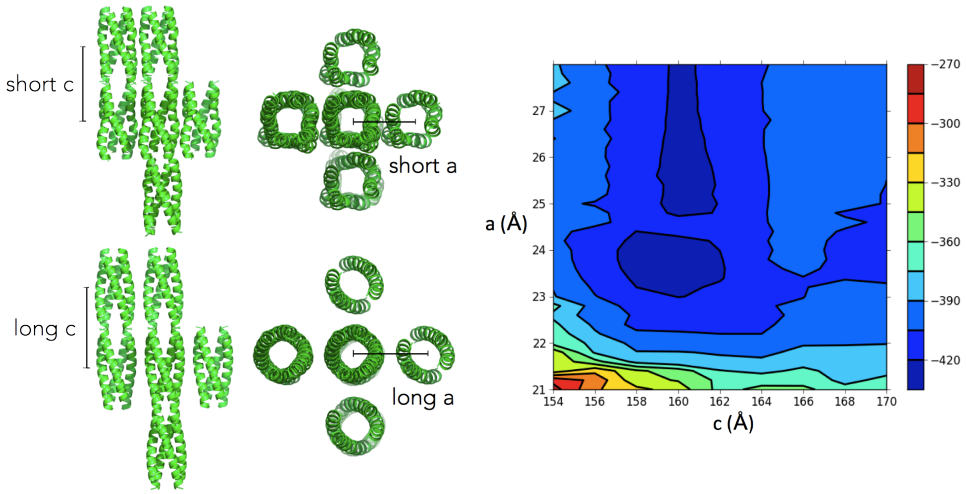


Figure 3.4. On the left, a representation of the degrees of freedom taken into account during the design. On the right, a contour map showing crystal configuration stability as a function of a and c edges of the cell in Rosetta Energy units.

Energy in terms of Rosetta score function units was evaluated for each candidate backbone. Best crystalline parameters found in our

search were selected to further sequence design[17]. A pool of sequences was derived, and two of them were chosen for the synthesis and characterization.

During the work to assess the designability of the selected crystal topology, a PDB search resulted in the finding of a tetrameric mutant of the GCN4, which crystallized in a similar topology, but without the intended polarity of the tubular substructures, which are antiparallel[18] (PDB ID: 2ccn). This coiled coil presented a slightly different superhelical frequency such that it accommodates a 33 residue long peptide in each layer of the crystal. This finding promoted the design of two more sequences of 33 residues whose design is based on the backbone of the GCN4 mutant.

3.1.2 Synthesis

In collaboration with the DeGrado Lab in the University of California, San Francisco, two QF analogues were synthesized based on the GCN4 design. Namely QF_6 and QF_33:

```
                g abcdefg abcdefg abcdefg abcdefg abcd
QF_6   001 Ace R LKQIIDR LEEIISD EYHHTNE LERIRKL LAER NH2 033
QF_33  001 Ace L LDEIKKW LEKIKED EKRHTQL LEKIKKI LEKI NH2 033
```

Furthermore, two sequences from *de novo* crystal design were synthesized:

```
                   defg abcdefg abcdefg abcdefg ab
W45Y27001 Ace IVYL LDEIKAD EKRHTQE LEWIAKQ LL NH2 027
W45Y28001 Ace IAEY LEKIEAD EKRHTQE LEWIRKQ LD NH2 027
```

Peptides were synthesized by SPPS with standard Fmoc protocols. The crude peptides were purified to homogeneity by preparative RP-HPLC. All the peptides were highly soluble in water (>20mg/mL). All the sequences present leucines and isoleucines in a and d positions, this motif is known to induce coiled-coil tetramerization even though anti-parallel bundle formation is directed by a well organized interaction pattern between the helix faces[9,19].

3.1.3 Spectroscopic characterization

3.1.3.a NMR spectroscopy

All the sequences have been analyzed via NMR. Interestingly QF_6 shows a well-resolved spectrum even in the apo- form, with uniform intensity amide protons ranging from 7.0 to 8.7 ppm (Figure 3.5). Slight changes were observed upon metal addition, indicating that the binding further stabilize the global folding.

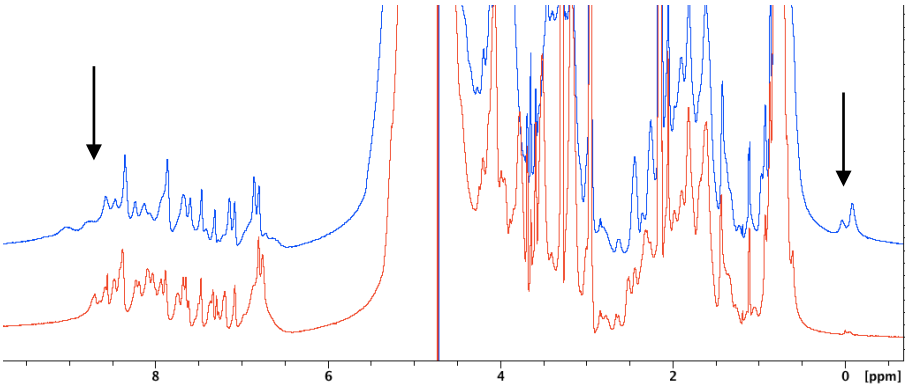


Figure 3.5. Apo (red line) and Zn bound (blue line) QF_6 NMR spectrum in 50mM phosphate buffer pH 6.5.

While QF_6 present a globular folding in solution, QF_33 is not detectable at neutral pH by solution NMR, probably because of a very slow rotational motion. This finding suggests the presence of an oligomerization state higher than four. This may indicate that the final purpose of protein crystallization is achievable in the tested designable space, even though the nucleation rate was too high. Figure 3.6 shows the spectrum of apo-QF_33 at acidic pH, showing a poor dispersed spectrum. Upon addition of NaOH, the sample undergoes dramatic signal drop.

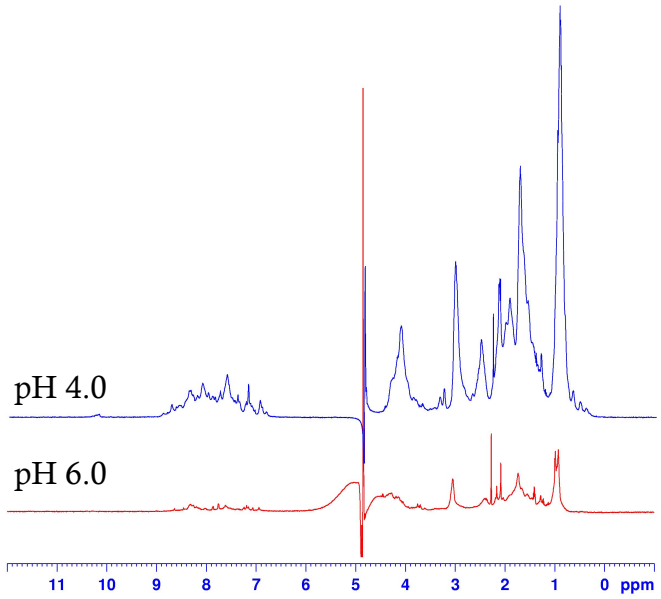


Figure 3.6. QF_33 ^{1}H-NMR spectrum in water at pH 4.0 (blue) and after addition of NaOH (red) to a final pH of 6.0

This finding gives us some hints for the design of successive sequences. In particular, QF_33 presents several hydrophobic residues

in ‘g’ positions, which may result in a hydrophobic patch suitable for bundle-bundle interaction.

Binding of zinc(II) was not detectable by NMR for the W45Y27 analogue, which showed a very low dispersion of amide peaks, demonstrating no organization at neutral pH and in presence of the metal (data not shown).

W45Y28 shows increasing structure content upon Zn(II) addition. ^{1}H-NMR spectrum in the amide region reaches a fine dispersed profile when a metal:peptide ratio of 0.5 is reached, demonstrating that the four-helix bundle is completely formed, upon binding of two metal ions (Figure 3.7).

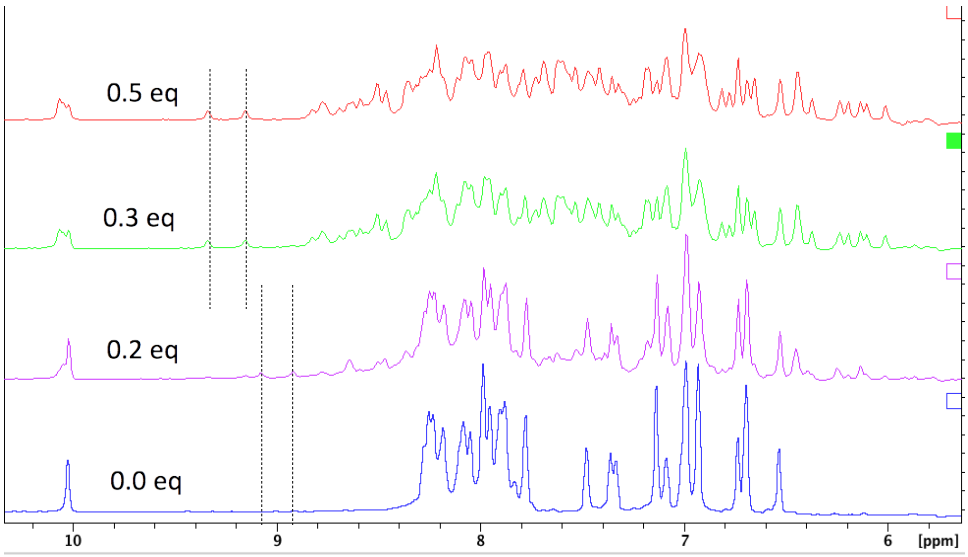


Figure 3.7. 0.3 mM W45Y28 NMR spectrum in the amide region at increasing metal to peptide ratio. Amide signals are well resolved and highly dispersed from 6 to 9 ppm.

At high fields for methyl protons, the spectral feature confirms the creation of a hydrophobic core (Figure 3.8).

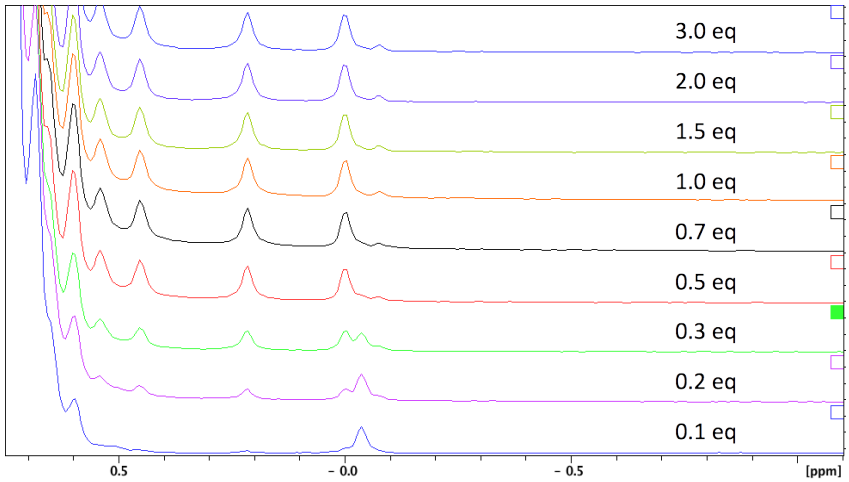


Figure 3.8. Aliphatic region showing the effect of water segregation on methyl signals. Definitive spectral features are reached at 0.5 equivalents (red).

A single low field peak for histidine involved in metal binding Hε2 proton, as previously assigned for DF compounds[13,20], suggests a D_2 symmetric environment, as intended in the design (Figure 3.9).

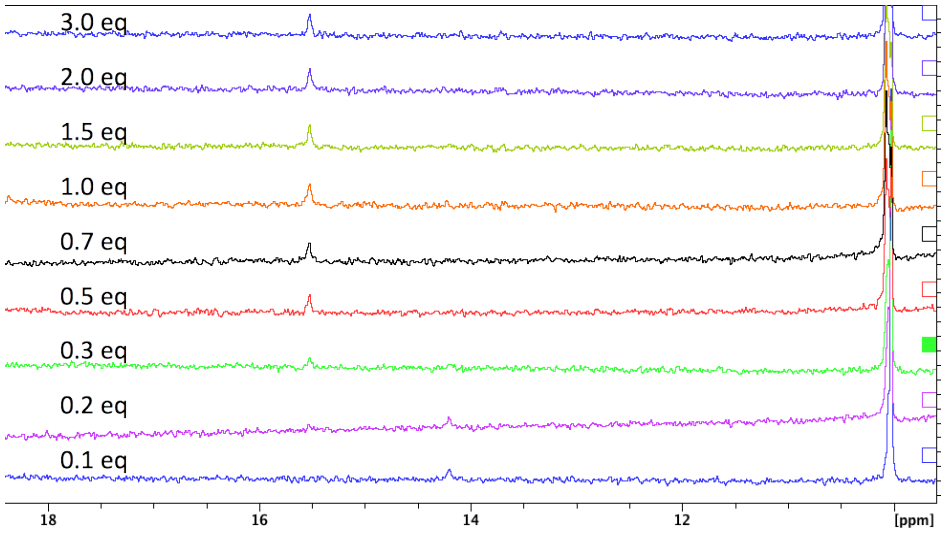


Figure 3.9. Lower field shift of histidine H ε 2 proton is due to H-bond designed interaction with second shell aspartic 11.

Even though no final evidence supports the binding of more than two metal ions, few more spectral features, appearing over 0.5 equivalents, corroborate the idea of the formation of a zinc cluster, as pointed out by NMR analysis of the partially deuterated peptide (Figure 3.10).

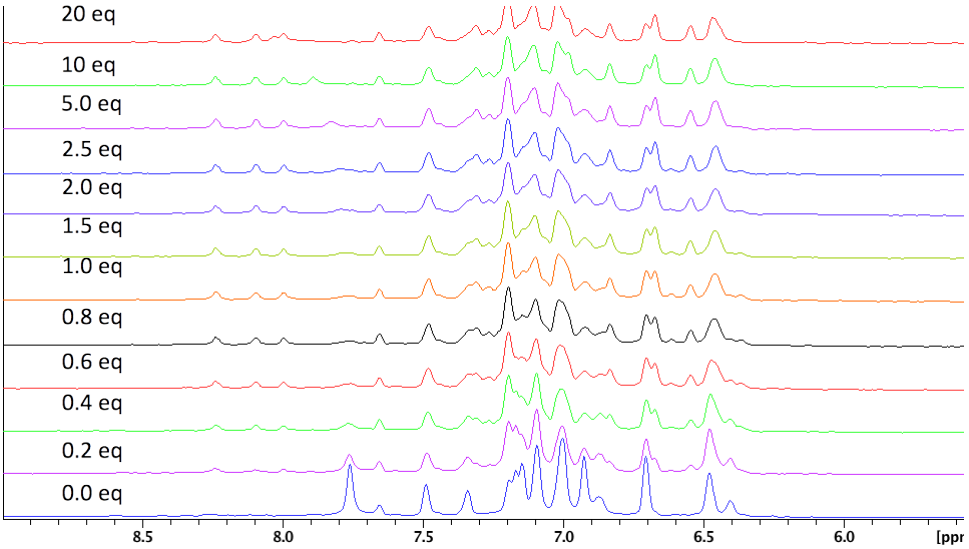


Figure 3.10. Slight changes in the aromatics region are still visible upon addition of Zn(II) over the 0.5 metal: peptide ratio.

3.1.3.b X-ray crystallography

QF_6 Apo protein formed good quality crystals for structure determination. Analysis of the diffraction pattern revealed a high-resolution structure (2.0 Å), with the desired $P4_1$ symmetry. We found an antiparallel four-helix bundle in the asymmetrical unit, with an almost perfect non-crystallographic D_2 symmetry. However, the sequence was not designed to obtain the correct polarization of the crystalline environment and it doesn't show the topology discussed

above. Tetramers are on top of each other along the coiled coil axis, as in the crystal design, but in one of the perpendicular directions they are rotated by 90° (Figure 3.11).

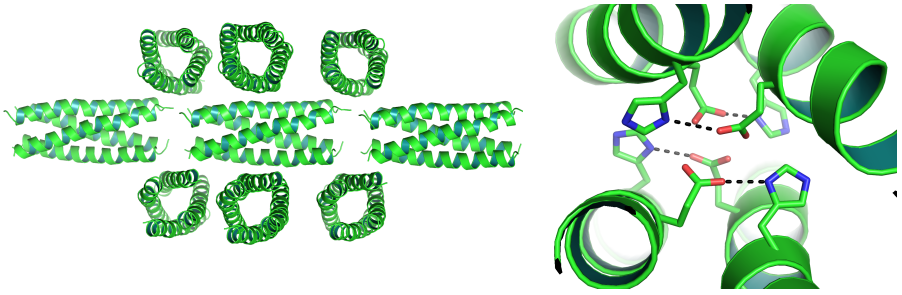


Figure 3.11. On the left, cartoon representation of Apo-QF_6 crystal topology, showing alternating orientations of the bundles (4_1 screw). Only nine bundles are shown for clarity. On the right, H-bond pattern of the site in the absence of metals bound. A dashed line corresponds to a H-bond interaction.

As predicted by the NMR analysis, QF_6 is folded even in the absence of the metals.

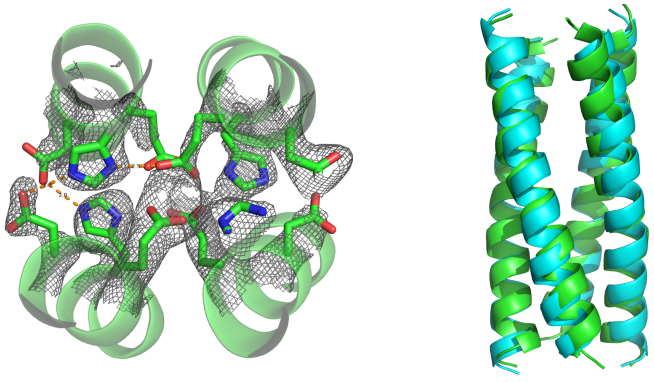


Figure 3.12. 2Fo-Fc Electron density map of the active site with the designed secondary interactions on the left (contour levels are 1.5σ). Cartoon representation of the superimposed crystal (green) and design (light blue) models on the right (1.3Å RMSD).

Only slight shifts of the carboxylate ligands allow H-bonding to His and water, rather than the metal. This stabilizes the energetic penalty associated with the burial of polar residues in the hydrophobic core (Figure 3.12, left). The designed and the crystallographic models deviate only by 1.3 Å RMSD, over the Cα trace, while a slightly higher 1.7 Å, for the all-atom superimposition, is observed.

Sequences designed for the correct crystal topology formed crystals in few hours within the conditions tested in a first screening, and data acquisition is now in progress.

3.2 Experimental section

Computational design

Design of the QF core unit was accomplished via a stepwise process by a combination of PyMOL and Accelrys Discovery Studio software. DF1 coordinates were submitted to the CCCP server (http://www.grigoryanlab.org/cccp/) in order to fit the backbone towards the generalized Crick model for coiled coils[9]. A global D_2 symmetric bundle was imposed, during the fitting, to enhance the convergence of the algorithm. Side-chain packing of the core residues was accomplished *via* dead-end elimination theorem, as implemented in Rosetta 3.5[21,22]. D_2 symmetry was retained *via* the Rosetta symmetry engine[23]. A command of this form was adopted for the side-chain packing:

```
fixbb.linuxgccrelease -in::file::s input.pdb -
symmetry:symmetry_definition D2.symm @../fixbb.flags
```

where flags adopted for this round were:

```
-database /PATH/TO/rosetta_database
-residues:patch_selectors CTERM_AMIDATION
-residues:patch_selectors ACETYLATED_NTERMINUS
-in:Ntermini ABCD
-resfile resfile_Core
-multi_cool_annealer 10
-packing:soft_rep_design
-ex1
-ex2
-nstruct 25
-overwrite
```

Resfile here permits to pack only apolar amino acids LIFVA, in *a* and *d* positions. Residues involved in the metal binding in the first and second coordination shells were kept fixed. At this stage, it is only needed a stable core for the next round of design.

The four-helix bundle was then minimized in Discovery studio without constraints, typed with charmm22 forcefield parameters, and minimized till convergence of the gradient, with a combination of steepest descent and conjugate gradient algorithms implemented in Discovery studio. Solvation energies were introduced *via* a Generalized Born implicit solvent algorithm[24]. Spherical cutoffs were adopted for long-range interactions, with an exponential switch from 1.4 nm to 1.6 nm.

Crystal design was easily implemented in Rosetta 3.5, thanks to the symmetry engine supported. Generation of the arrays of candidate backbones, as function of the **a** and **c** edges of the $P4_1$ unit cell, was accomplished by iterating these parameters in the generation of Rosetta symmetry definition files:

```
make_symmdef_file.pl -m CRYST -p input.pdb -c ITERa ITERa ITERc
90.0 90.0 90.0 -s P 41 > /ITERc_ITERa_P41.symm
```

Where ITERa and ITERc were simply iterated with the desired values, as described above in the results. Rosetta score was evaluated for each **a** and **c** pair and plotted with GNUPLOT[25]. Once the best set of cell edges was found, the design of the sequences was made, as described earlier for the core unit. However, in this case all interface residues were free to mutate to all the amino acids, but Met, Cys, His for synthetic and metal binding reasons. Tryptophan and tyrosine, both for 280 nm absorption and crystal packing purposes, were imposed in some positions resulted from a residue scan in $P4_1$ crystal environment. Sequences obtained are classified by which aromatic residue they show and in which position. Following the numeration of DF1 helix 2, possible combinations are: W45WY27, W45WY28, W45WY49, Y41WY27, Y41WY28, Y41WY44, Y41WY49. Sequences were then manually inspected and evaluated in terms of number of interactions, in the crystal and voids in the packing, and three sequences were selected for synthesis. Only W45Y27 and W45WY28 sequences described in the results showed some hits in the

first crystal screen. In this process, coordinates of the residues involved in the metal binding were always kept fixed.

Designs derived from GCN4 mutant were obtained by simple insertion of metal binding and second shell residues onto the 2ccn derived scaffold, followed by repacking with D_2 symmetry definition file in Rosetta 3. The two sequences selected for synthesis were chosen by manual inspection of the computationally derived sequences.

Peptide synthesis

QF_33 was synthesized using solid phase synthesis with standard Fmoc protocols on 0.1 mmol scale, using an Fmoc-PAL-PEG PS resin (with a substitution level 0.20 mmol/g).

The protected amino acids used in the synthesis of QF_33 were the following:

Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Glu(OtBu).OH, Fmoc-Leu-OH, Fmoc-Gln(Trt)-OH, Fmoc-Thr(tBu)-OH, Fmoc-His(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Trp(Boc)-OH.

The peptide cleavage from the resin and side chain deprotection was carried out using a solution of 95% TFA, 2.5% TIS, 2.5% H_2O (v/v/v 20 mL).

The crude product was analyzed to homogeneity by analytical reverse-phase high-performance liquid chromatography (RP-HPLC), performed with a Vydac C18 column (4.6 mm x 150 mm; 5 μm), eluted with an H_2O 0.1% TFA (solvent A) and CH_3CN 0.1% TFA (solvent B) linear gradient, from 10 to 80% (solvent B) over 70 min, at

1 mLmin^{-1} flow rate, in all analysis. The chromatogram shows a main peak at R_T = 53.014 min. The identity of QF_33 has been confirmed by mass spectrometry (theoretical mass: 4171,037 Daltons; observed mass: 4171.0 ± 0.5 Daltons). See Figure 3.13.

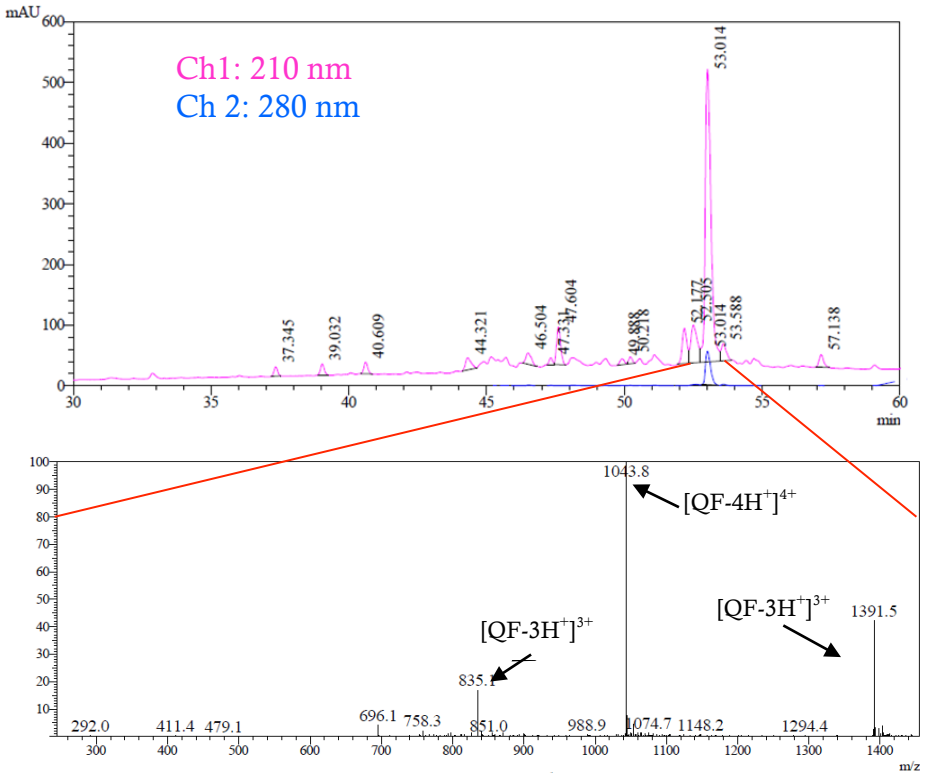


Figure 3.13. RP-HPLC-MS chromatogram of crude QF_33 peptide

Peptide was further purified to homogeneity, by preparative RP-HPLC to a single, symmetric peak. The columns used were a Vydac C18 column (1 cm x 25 cm; 5 μm), at a flow rate of 5 mL min^{-1} and Vydac C18 column (2.2 cm x 25 cm; 10 μm), at a flow rate of 21 mL min^{-1} with a gradient of acetonitrile in 0.1% aqueous TFA, from 30 to 56% over 20 min (Figure 3.14).

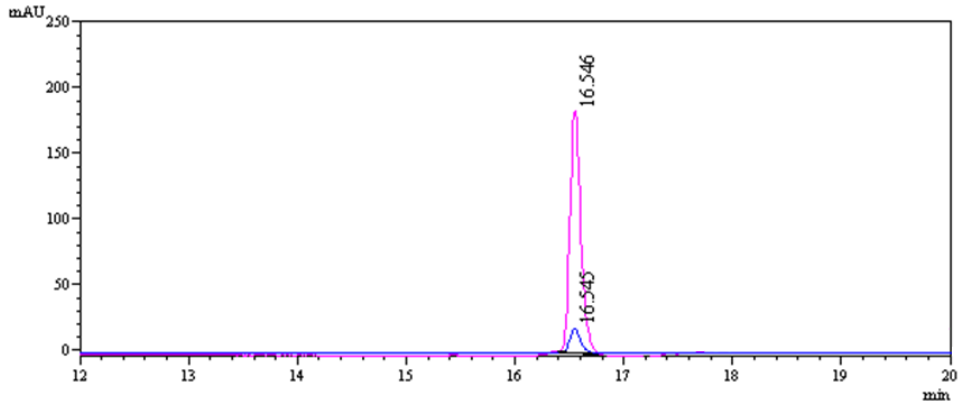


Figure 3.14. RP-HPLC chromatogram of pure QF_33 peptide

All other peptides were synthesized on a Symphony automatic synthesizer in a 0.2 mmol scale, with standard Fmoc protocols, as previously reported[26]. Peptide purification was achieved via RP-HPLC with methods equivalent to what has been previously described for QF_33.

NMR Experiments

NMR spectroscopy was used to analyze the all the QF analogues. All solutions for NMR spectra were prepared with 10 % of D_2O. The apo form was analyzed at concentration of 500 μM in H_2O, for QF_33, and 300 μM in 10 mM NaP buffer pH 6.5, for all the other peptides. Small aliquots, from properly concentrated aqueous $ZnCl_2$ solutions, were added *in situ*. Metal concentration was ascertained *via* ICP-MS methods. NMR spectra of QF_33 were acquired at 25 °C, on a Bruker Avance 600 spectrometer, equipped with a triple resonance cryo-probe. NMR spectra of the other analogues were acquired at 25

°C, on a Bruker Avance 900 spectrometer. Suppression of the water signal was accomplished by excitation sculpting sequence.

X-Ray Crystallography

Crystal structure of QF_6 is a gentle contribution to this work by Jenny Hu, PhD University of California San Francisco. The stock protein solution was prepared by dissolving 2 mg of protein in 50 μl of H_2O, followed by the addition of 10 eq of aqueous zinc acetate. Small crystals were obtained from hanging drops at 25°C from 0.2 M sodium thiocyanate, 18% PEG3350, 0.1M MES, pH 6.0. A single crystal was frozen in liquid nitrogen and used for data collection, under a stream of cold nitrogen (at 100 K), by using glycerol as cryoprotectant. Data were collected by using an ADSC Quantum 315r detector at the synchrotron ALS Beamline 8.3.1 (Berkeley, CA, USA). Data were reduced by using MOSFLM in the $P4_1$ space group with cell dimensions of a=b 37.557 Å, c 84.854 Å. Resolution 37.5~2.1 Å; R/R_{free} 0.24/0.26 so far. Structure is still under the final refinement process.

References

(1) Kharenko, O. A.; Kennedy, D. C.; Demeler, B.; Maroney, M. J.; Ogawa, M. Y. *J. Am. Chem. Soc.* **2005**, *127*, 7678.

(2) Xie, F.; Sutherland, D. E. K.; Stillman, M. J.; Ogawa, M. Y. *J. Inorg. Biochem.* **2010**, *104*, 261.

(3) Grzyb, J.; Xu, F.; Weiner, L.; Reijerse, E. J.; Lubitz, W.; Nanda, V.; Noy, D. *Biochim. Biophys. Acta BBA - Bioenerg.* **2010**, *1797*, 406.

(4) Wu, S.; Bellei, M.; Mansy, S. S.; Battistuzzi, G.; Sola, M.; Cowan, J. A. *J. Inorg. Biochem.* **2011**, *105*, 806.
(5) Lubitz, W.; Reijerse, E. J.; Messinger, J. *Energy Environ. Sci.* **2008**, *1*, 15.
(6) Howard, J. B.; Rees, D. C. *Chem. Rev.* **1996**, *96*, 2965.
(7) Ferreira, K. N.; Iverson, T. M.; Maghlaoui, K.; Barber, J.; Iwata, S. *Science* **2004**, *303*, 1831.
(8) Lanci, C. J.; MacDermaid, C. M.; Kang, S.; Acharya, R.; North, B.; Yang, X.; Qiu, X. J.; DeGrado, W. F.; Saven, J. G. *Proc. Natl. Acad. Sci.* **2012**, *109*, 7304.
(9) Grigoryan, G.; DeGrado, W. F. *J. Mol. Biol.* **2011**, *405*, 1079.
(10) Zhang, J.; Grigoryan, G. In *Methods in Enzymology*; Elsevier, 2013; Vol. 523, pp. 21–40.
(11) Helling, R.; Li, H.; Mélin, R.; Miller, J.; Wingreen, N.; Zeng, C.; Tang, C. *J. Mol. Graph. Model.* **2001**, *19*, 157.
(12) Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F. *Proc. Natl. Acad. Sci.* **2000**, *97*, 6298.
(13) Maglio, O.; Nastri, F.; Martin de Rosales, R. T.; Faiella, M.; Pavone, V.; DeGrado, W. F.; Lombardi, A. *Comptes Rendus Chim.* **2007**, *10*, 703.
(14) Nastri, F.; Bruni, R.; Maglio, O.; Lombardi, A. In *Coordination Chemistry in Protein Cages*; Ueno, T.; Watanabe, Y., Eds.; John Wiley & Sons, Inc., 2013; pp. 43–85.
(15) Adams, P. D.; Baker, D.; Brunger, A. T.; Das, R.; DiMaio, F.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C. *Annu. Rev. Biophys.* **2013**, *42*, 265.
(16) Kaufmann, K. W.; Lemmon, G. H.; DeLuca, S. L.; Sheehan, J. H.; Meiler, J. *Biochemistry (Mosc.)* **2010**, *49*, 2987.
(17) Li, Z.; Yang, Y.; Zhan, J.; Dai, L.; Zhou, Y. *Annu. Rev. Biophys.* **2013**, *42*, 315.
(18) Yadav, M. K.; Leman, L. J.; Price, D. J.; Brooks, C. L.; Stout, C. D.; Ghadiri, M. R. *Biochemistry (Mosc.)* **2006**, *45*, 4463.
(19) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401.

(20) Maglio, O.; Nastri, F.; Lombardi, A. In *Ionic Interactions in Natural and Synthetic Macromolecules*; Ciferri, A.; Perico, A., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012; pp. 361–450.
(21) Leaver-Fay, A.; Kuhlman, B.; Snoeyink, J. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **2005**, 16.
(22) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364.
(23) DiMaio, F.; Leaver-Fay, A.; Bradley, P.; Baker, D.; André, I. *PLoS ONE* **2011**, *6*, e20450.
(24) *J. Comput. Chem. 24*, 1691.
(25) Williams, T.; Kelley, C. *Off. Gnuplot Doc. Httpsourceforge Netprojectsgnuplot* **2010**.
(26) Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; Degrado, W. F. *Science* **2011**, *332*, 1071.

Chapter 4: *Conclusions*

4 Conclusion and perspectives

In this work we expanded the range of metal binding geometries accessible with the DF, a family of *de novo* four-helix bundle metalloproteins. We realized two new classes of DF-like proteins from first principles, by means of *de novo* design. Starting from the C_2 symmetry of the first member of the DF family, DF1, the design process was devoted towards a loss of symmetry, with the DF-Click on the one side, and an increased D_2 symmetry with the QF on the other.

4.1 DF-Click: Building a library of new compounds and functions

The DF-Click design is an asymmetric metalloprotein, which takes the positive aspects of its asymmetric precursors DF_{Tet} and DF_{SC}. We successfully developed a synthetic protocol for the covalent ligation of two different DF monomers, by means of Click chemistry. The protocol potentially gives the opportunity to generate a whole library of ligands and we have shown its selectivity towards well-behaved dimers, as predicted by the design. We synthesized and characterized the first member of this subset of compounds, named DF-Click1. Spectroscopic analysis demonstrates the global folding and the asymmetry of the metal binding site, thus confirming the correctness of the design. It binds zinc(II) and iron(II) with a greater selectivity with respect to other DFs, showing that the fine-tuning of the secondary

shell coordination influences the metal binding properties. In perspective, DF-Click strategy will be adopted in the generation of compounds with new catalytic activities, which were not accessible by the previous DFs. These compounds may have an immediate impact on catalysis, biosensing and mild oxidation of organic water pollutants.

4.2 QF: New generation nanomaterials for light harvesting

The QF design is a D_2 symmetric tetranuclear metalloprotein. We adopted an innovative approach for the *de novo* design of a protein crystal. We also developed an automated protocol for crystal design, which is compatible with the widely implemented Rosetta suite. Our design process led to few sequences, which show the intended folding and are able to bind the metal cofactor. Sequences corresponding to the crystal design formed crystals within hours, and collection of the diffraction data tested the correctness of the design protocol. In perspective, we aim to develop new nanocomposite materials, based on the self-assembly of metal binding polypeptide chains, which may reveal unprecedented features in the fields of bioengineering and energy remediation.

List of abbreviations

Abs: Absorbance
Ac: Acetyl
Ac_2O: Acetic anhydride
ACP: stearoyl-acyl carrier protein
Ala: Alanine
Arg: Arginine
Asn: Asparagine
Asp: Aspartic acid
AOX: alternative oxidase
BMMs: bacterial multicomponent monooxygenases
Boc: t-Butoxycarbonyl
CaM: calmodulin
CATH: class, architecture, topology, homology
CCCP: Coiled Coil Crick Parameterization
CD: Circular dichroism
CuAAC: Cu(I)-catalyzed azide-alkyne 1,3-dipolar cycloaddition
Cys: Cysteine
DCM: Dichloromethane
DEE: dead-end elimination
DF: Due Ferro
DOSY: Diffusion-ordered spectroscopy
DIEA: Diisopropylethylamine
DMF: Dimethylformamide
DNA: deoxyribonucleic acid
EDT: Ethanedithiol
EPR: electron paramagnetic resonance spectroscopy
ESI-MS: Electrospray ionization mass spectrometry
EXAFS: Extended X-Ray Absorption Fine Structure
FAD: flavine adenine dinucleotide
FID: free induction decade
Fmoc: 9-Fluorenylmethoxycarbonyl
Gln: Glutamine
Glu: Glutamic acid

Gly: Glycine
HATU: N-[(dimethylamino)-1H-1,2,3-triazolo[4,5-b]pyridin-1-yl-methylene]-N methylmethanaminium hexafluorophosphate N-oxide
HCTU: O-(1H-6-Clorobenzotriazole-1-yl)-1,1,3,3-tetrametiluronium hexafluorophosphate
HEPES: 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His: Histidine
HOBt: N-hydroxybenzotriazole
Hr: hemerythrin
ICP-MS: Inductively Coupled Plasma Mass Spectrometry
Ig: immunoglobulin
Ile: Isoleucine
LC-MS : liquid chromatpgraphy mass spectrometry
Leu: Leucine
Lys: Lysine
LMCT : Ligand to Metal Charge Transfer
MaDCaT: Mapping of Distances for the Categorization of Topology
MeOH: Methanol
sMMO: soluble methane monooxygenase
Mmt: Methoxytrityl
MW : molecular weight
NCAA : non-canonical amino acid
NMP: N-Methyl-2-pyrrolidone
NMR: Nuclear magnetic resonance
NOE : Nuclear Overhauser Effect
OEC : oxygen evolved complex
PAL : 5-[3,5-Dimethoxy-4-(aminomethyl)phenoxy]pentanoic acid
Pbf: 2,2,4,6,7-Pentamethyldihydrobenzofurane-5-sulfonyl
PDB: Protein Data Bank
PEG : poly-ethilene glycole
PS : poly-styrene
PTOX: plastid terminal oxidase
QF: Quattro ferro
QM-MM: quantum mechanics/molecular mechanics
SCADS: statistical computationally assisted design strategy
SCOP: structural classification of proteins
Ser: Serine

SOD: superocide dismutase
SPPS: solid-phase peptide synthesis
RMSD: root mean square deviation
RNR: ribonucleotide reductase
RP-HPLC: reversed-phase high performance liquid chromatography
RSM: recurrent structural motif
t-Bu: tert-butyl
TCEP: Tris(2-carboxyethyl)phosphine hydrochloride
TFA: Trifluoroacetic acid
Thr: threonine
TIM: triosephosphate isomerase
TIS: triisopropyl silane
TLC: Thin Layer Chromatography
TMO: Toluene monooxygenase
T3MO/T4MO: toluene-3/4-monooxygenase
ToMO: toluene *o*-xylene monooxygenase
Trt: Trityl
Tyr: Tyrosine
UV-Vis: UV-Visible

PUBBLICATIONS AND COMMUNICATIONS

Chino M. et al., "Click Chemistry, a useful strategy for developing helix-loop-helix heterodimer, housing asymmetric di-metal binding sites", manuscript in preparation.

CHINO M., Nastri F., Maglio O., Pavone V., Lombardi A.

XL Congresso Nazionale della Divisione di Chimica Inorganica della Società Chimica Italiana. Sestri Levante (GE), 9-13 September 2012, p. OC5 (Oral Presentation)

Title. *Four-helix bundle scaffolds housing metal binding sites*

Abstract. Nature has selected very efficient metal binding scaffolds by evolution, thus reaching a plethora of functions adopting only few ligands, slightly different geometries and most easily available metal ions. In this regard, four-helix bundle is one of the most recursive scaffold in nature. It is very stable, fine tunable, and most of all "designable". In this contest, we previously developed, through an iterative design process, C_2 symmetric four helical bundles (DFs) that contain dicarboxylate bridged di-nuclear metal centers.

Aimed to obtain asymmetric di-metal binding site, our design is progressing towards hetero-dimeric models, composed by two different helix-loop-helix monomers. We adopted two complementary approaches in order to address the hetero-dimer formation: covalent binding, through Copper catalyzed Azide-Alkyne cycloaddition (CuAAC), or self-assembly, through properly designed non-covalent interactions. Stability and metal binding properties of the designed molecules are being analyzed through circular dichroism and UV-vis spectroscopies.

Further on, a multi-step design process brought to a four-helix bundle tetramer, namely QF (Quattro Ferri, Italian for Four Iron),

aimed to house up to four metal ions. QF will allow the exploration of different activities and possibly widespread applications in analyte sensing, electron exchange/storage, solar energy conversion. In summary, the adopted design process is broadening our "designable" space, from *pseudo*-D_2 configurations, as found in natural proteins, to higher symmetric D_2 four-helix bundles, obtained through the self-assembly of a single helix.

CHINO M., Nastri F., Maglio O., Pavone V., Lombardi A.

XXXVII "A. Corbella" Summer School. Gargnano (BS), June 18-22 2012, p. 1

Title. *iTOL-DF: developing new di-iron containing proteins as catalysts for oxidation reactions.*

Abstract: Asymmetric catalytic sites are chosen by nature to obtain complex, chemoselective and specific functions. In this regard, toluene monooxygenase hydroxylase (ToMOH) is a perfect example, as it is able to hydroxylate a variety of organic compounds, such as halogen alkanes and aromatic derivatives, consuming equivalents of dioxygen or hydrogen peroxide. This work aims to design new di-metal containing four-helix bundle scaffold devoted to mimic ToMOH active site geometries and second shell interactions. Starting from previous results of the DF-series, we designed a new model, iTOL-DF, which is composed by two helix-loop-helix monomers, covalently bound through side-chain modifications. Standard Fmoc-SPPS was used to synthesize the two monomers, which were further conjugated through Copper catalyzed Azide-Alkyne cycloaddition (CuAAC). Synthetic steps were followed by RP-HPLC and the desired product was identified through mass spectrometry. Stability and metal binding properties are being analyzed through circular dichroism and UV-vis spectroscopies. The results so far obtained are very promising for the

application of low molecular weight artificial enzymes in organic pollutant biosensing.

Ringhieri P., Nastri F., Bruni R., CHINO M., Maglio O., Lombardi A., Pavone V.

XXIV Congresso Nazionale della Società Chimica Italiana, Lecce 11-16 Settembre 2011

Title. *Synthetic transducers in affinity third generation electrochemical biosensors.*

Abstract. A biosensor is an analytical device that relies on a biological recognition element, communicating with a signal transducer that provides a measurable response related to the concentration of the analyte. Electrochemical biosensors, which exploit an electron transfer (ET) event at a solid electrode for signal transduction, are the most promising biosensors, in terms of ease of construction, costs, versatility and miniaturization. In particular, third generation biosensors exploit the direct electron transfer occurring between an electrode and a redox-active species immobilized onto it. The biological interaction is transduced into a current signal, thus resulting in an amperometric detection of the analyte.

Efficient direct ET reactions have been reported only for a restricted number of redox enzymes; several studies based on heme-proteins adsorbed or immobilized on various electrodes showed that high-molecular weight enzymes are often not suitable for direct electrical communication with the electrode. Our strategy to overcome these limitations is the use of artificial low-molecular weight proteins, designed on rational bases, to own the required activity. We developed a class of covalent heme-peptide conjugate, named Mimochromes, with the aim of understanding the effects of the peptide chain composition and conformation in modulating the heme redox potential.

An affinity electrochemical immunosensor was developed using Mimochromes as redox tags for the detection of the binding events. The artificial ET protein, functionalized with a suitable recognition element, was covalently anchored on a gold electrode through self-assembled monolayers. The biosensor was characterized with voltammetric techniques and tested for analyte recognition.

CHINO M., Nastri F., Maglio O., Lombardi A., Pavone V.

XXIV Congresso Nazionale della Società Chimica Italiana, Lecce 11-16 Settembre 2011

Title. *Developing new di-iron containing proteins as catalysts for oxidation reactions.*

Abstract. The quantitative and high efficiency detection of several analytes is required in many different areas. Since the basic concept of enzymatic electrode by Clark and Lyons and after successful applications of glucose-oxidase based biosensors in blood glucose measurement, a bio-revolution in chemical analysis has been occurring. Recently, interest in organic pollutant biosensing has been growing. In this regard, Toluene Monooxygenase family (TMOs), a four components bacterial multi-component monooxygenase, can be of crucial utility as it is able to hydroxylate a variety of organic compounds, such as halogen alkanes and aromatic derivatives, consuming equivalents of dioxygen or hydrogen peroxide. A cost/time effective alternative to the classic microbial biosensor is the adoption of synthetic compounds able to mimic TMO activity, bypassing enzyme purification process and overtaking cell diffusion limits.

Full-organic ligands coordinating two metal centers, which are able to mimic enzyme reactivity, and, small peptides, which are able to fold into four-helix bundle motives, have been developed in the last decades. An interesting strategy is the synthesis of small peptide chains, with amino acid sequence containing information both for a

proper folding and for metal binding. In the last years, several models of a di-metal self-assembling four-helix bundle motif has been obtained by means of de novo design, the DF series. Recently, redox activity has been achieved with the di-iron DF3 model. It is able to bind and successfully oxidize phenolic substrates, similar to the natural enzyme alternative oxidase. This result assures the suitability of DF-like scaffolds for engineering new catalytic site through various metals and coordination geometries. Here, we report the design and a preliminary computational study of new DF variants, mimicking TMO active site geometries and second shell interactions.

CHINO M., Andreozzi C., Maglio O., Nastri F., Pavone V., Lombardi A.

Scuola Nazionale di Chimica Bioinorganica, Siena 3-6 Luglio 2011

Title. *Four-helix bundle biocatalysts as novel biosensing elements.*

Abstract. The quantitative and high efficiency detection of several analytes is required in many different areas. Since the basic concept of enzymatic electrode by Clark and Lyons and after successful applications of glucose-oxidase based biosensors in blood glucose measurement, a bio-revolution in chemical analysis has been occurring.

A biosensor is a device that detects, transmits and records information regarding a physiological or biochemical change. It is composed by a biological component, which is deputed for recognition and detection, an immobilizing component, which localizes the enzymes near the previous component, an electronic transducer, which converts a biochemical signal into an electrical response.

Biosensors are replacing the complex and time-consuming laboratory-based methods, used so far. They offer the advantages of providing rapid, sensitive, inexpensive analytical results, as well as operating directly, selectively, and continuously. The main reason why

biosensing has not overwhelmed other methods resides in the enzyme isolation and purification. In order to avoid the laborious and expensive operations of enzyme purification the use of whole viable or non-viable cells has been proposed. A different strategy uses biomimetic compounds, which are simultaneously cost-effective, while retaining enzymatic pros of selectivity and real-time response.

The four-helix bundle is a very common motif in nature, well suited for the development of biomimetic compounds. It is composed by four α-helices packed against each other roughly lengthwise. This motif is encountered in numerous and functionally diverse proteins, such as interleukin-2, apolipoprotein, human growth hormone, as well as in metalloproteins. In fact, the four-helix bundle can efficiently bind metal cofactors introducing a great variety of functions, simply modulating metal-protein matrix interactions. Representative examples are carboxylate bridged-diiron proteins, which serve as dioxygen activator (bacterial multicomponent monooxygenases - BMMs, ribonucleotide reductase R2 subunit, stearoyl-Acyl Carrier Protein Δ^9-desaturase); dioxygen binding (hemerythrin and myohemerytrin), ferroxidases (ferritins, bacterioferritins).

Full-organic ligands coordinating two metal centers, which are able to mimic enzyme reactivity, and, small peptides, which are able to fold into four-helix bundle motives, have been developed in the last decades. An interesting strategy is the synthesis of small peptide chains, with amino acid sequence containing information both for proper folding and for metal binding. By means of *de novo design,* our group developed di-iron-oxo-protein models (named DFs). They are made up of two 48-residue *helix-loop-helix* (a_2) motifs, able to specifically self-assemble into an antiparallel four-helix bundle [6-8]. Each monomer of the DF dimer contains the well-known Glu_a-Xxx_b-Xxx_c-His_d binding motif and a single Glu in an "a" position of the metal-binding heptad. Through iterative cycles of design and characterization, the coordination and functional properties of the initial model were improved, and recently, redox activity has been achieved with the di-iron DF3 model. It is able to bind and

successfully oxidize phenolic substrates, similar to the natural enzymes alternative oxidase (AOX) and plastid terminal oxidase (PTOX), reducing dioxygen. This result assures the suitability of DF-like scaffolds for engineering new catalytic site, containing various metals and coordination environments.

In this perspective, we are adopting DF-scaffolds in order to obtain several novel peptide components for high-performing biomimetic enzyme based sensors. Our interest mainly focuses on the biosensing of organic pollutants. Therefore, Toluene Monooxygenase family (TMOs), a four components BMM, may result of crucial utility, as it is able to hydroxylate a variety of organic compounds, such as halogen alkanes and aromatic derivatives (i.e. benzene, toluene, ethyl-benzene), consuming equivalents of hydrogen peroxide.

Our efforts in engineering a di-iron site with monooxygenase activity in the DF scaffold will be presented. The use of the newly developed molecules as active component of biosensors will also be discussed.

ACKNOWLEDGEMENTS

Se riuscissi a citare tutte le persone che hanno contribuito alla riuscita, voglio dire miracolosa, di questo lavoro di tesi, potrebbero volerci molte più pagine di quanto non sia la tesi stessa. Mi scuso, dunque, con tutti coloro che non citerò, ma ai quali riconosco un ruolo importante nel percorso che mi ha portato qui; siete stati importanti per me come l'acqua per il raccolto.

Spesso è difficile scegliere come e con chi cominciare, e spesso si confonde il primo ringraziamento come un atto di cortesia, quasi dovuto. In questo caso non c'è nessun dubbio, il mio primo e profondamente sentito ringraziamento va alla Prof. Angela Lombardi, coautrice di questo lavoro. Lei è stata per me più di un tutor, è stata il mio primo mentore in questi tre anni costellati da mille peripezie. Mi ha sempre spinto a fare del mio meglio e a dare sempre di più, e spero tanto di averla ripagata almeno in parte dell'investimento fatto su di me, nonostante tutti i miei grandi limiti e il mio disordine. La passione che ho maturato per quest'argomento la devo a lei, e le immense opportunità di crescita professionale che mi ha dato saranno sempre un cavalleresco debito tra scienziati che avrò sempre nei suoi confronti.

Il Professore Vincenzo Pavone è stato indiscutibilmente nel bene e nel male il mio primo motivatore. Non è stato facile interfacciarsi l'uno all'altro, ma chi mi conosce sa l'enorme rispetto che provo nei suoi confronti e quanto questo mi limiti nel rapporto con lui per via della mia deprecabile difficoltà ad accettare l'autorità, che mi ha portato sin da piccolo a mettermi sempre all'opposizione. Sto

provando a crescere, e guardandomi indietro ora ci sono state tante volte che ho seguito i suoi consigli, più di quanti lui stesso si aspetterebbe da me e che per indole non ho mai pubblicizzato. Se un domani dovessi essere uno scienziato, anche solo la metà di quanto lo è il professore, mi sentirei grato della strada percorsa.

Alla Dott. Ornella Maglio la mia tutor devo gran parte del lavoro di risonanza magnetica. Quel poco che conosco di NMR, lo conosco grazie a lei. La pazienza che ha mostrato sempre nei miei confronti, mi ha aiutato a fare miei concetti che pensavo mi fossero preclusi all'inizio di questo periodo di dottorato. Da lei ho imparato ad amare la scienza e il proprio lavoro anche quelle volte che la vita ti si mette di traverso.

Se c'è qualcuno che non smetterò mai di ringraziare è la Prof. Flavia Nastri. Il suo aiuto nello sviluppo del metodo di Click e nell'analisi spettroscopica sono stati cruciali in questo lavoro di tesi. Dove tutti, compreso io, vedevano fallimenti, lei mi ha dato la chiave di lettura giusta per interpretare i miei dati. Se c'era qualcuno da "scocciare" con i problemi di laboratorio di tutti i giorni, questa era la professoressa Nastri. Da lei ho imparato che la pragmaticità si applica alla forma come alla sostanza.

Voglio ringraziare il mio supervisore, il Prof Piero Pucci. Ha saputo dirigere i miei passi in maniera discreta, corollando il mio lavoro con una serie di domande che mi hanno sempre portato in luoghi che altrimenti non avrei esplorato.

Quando si comincia a lavorare in questo campo non pensi che potrai mai avere la possibilità di conoscere i miti della chimica, quelli

dei libri, per intenderci. Io sono onorato di aver avuto la possibilità non solo di conoscere ma di lavorare braccio a braccio con Bill DeGrado. Il suo lavoro mi ha ispirato sempre, anche prima di conoscerlo. Quando poi conosci Bill, cambi completamente la prospettiva su quello che ritenevi impossibile. L'impossibile per Bill è solo un sogno che non pensi di meritare. Mi ha accolto nel mio periodo a San Francisco in una maniera per nulla scontata. Ho appurato che la sua grandezza si misura nell'umiltà con cui ci si pone in ogni rapporto, e che il rispetto non è una questione di numeri di citazioni o di regole, ma di visione, di ispirazione, di passione. Bill mi ha insegnato che il sapere ti apre gli occhi, ma il non sapere ti apre la mente.

Susan ha aperto la porta della sua casa e mi ha accolto come un figlio. Non saprò mai come ringraziarla. Ero un pesciolino di acqua dolce buttato nell'oceano, e lei mi ha dato le basi su cui partire per ricavarmi il mio spazio. Grazie per non avermi mai chiesto niente in cambio, la gratuità delle tue azioni rende gloria a Dio. Sono sicuro che ti ricolmerà di grazia.

Uno speciale ringraziamento va alle persone che hanno contribuito attivamente alla raccolta dei dati, e alla stesura di questa tesi. In particolar modo, voglio ringraziare le persone che ho seguito nel loro lavoro di tesi specialistica. Per meglio dire: le persone che mi hanno seguito durante il loro lavoro di tesi. Perché il loro aiuto giornaliero nel tenermi nelle righe mi aiuta a non disperdere completamente l'enorme quantità di lavoro che faccio. Sono state la mia ancora di salvezza, la mia memoria. Abbiamo imparato a fare squadra e a far sì

che i difetti dell'uno fossero nascosti dai pregi dell'altro, ed io sono stato molto nascosto. Spero, a mia volta, di aver regalato loro qualcosa di me.

Daniela il tuo contributo è stato totale nel lavoro di DF-Click. Se non fosse per l'aiuto che mi hai dato nella stesura della tesi, questo non sarebbe stato possibile. Mi auguro che tu possa sempre di più acquisire sicurezza in tutto quello che fai e ad accettare l'idea che meriti molto di più di quello che dai per scontato di meritare.

Rosa ha contribuito sostanzialmente nel lavoro di QF, aiutandomi con la sua tenacia a fare luce sulle più oscure questioni di questo impegnativo progetto. Avrei voluto più fortuna per il nostro lavoro su QF_33, ma penso che è nelle delusioni che si misuri il vero valore di uno scienziato, e sono felice di esserci passato con Rosa come compagna.

Violetta ed io ce la siamo vista brutta con la Click, al punto da rifiutare qualunque capo d'abbigliamento che non fosse munito di zip! Abbiamo tenuto duro, e l'abbiamo fatto insieme, raggiungendo importanti risultati per un novellino della sintesi organica come me. Sei stata una pioniera e ti ringrazio di esserti fidata di me, nonostante non sia altro che un inesperto scienziato, spesso pasticcione.

Poi come non menzionare le persone con cui ho condiviso, o tuttora condivido, l'onore (e perché no, l'onere) il laboratorio dell'AMEG.

Liliana, ho sempre visto in te un esempio da imitare, dal punto di vista scientifico e non. Non ho mai capito bene se ti piaccio oppure no.

So solo che qualunque cosa pensi, hai ragione. Liliana ha sempre ragione! Provare per credere.

Tina con te soffro ancora di una sindrome da distacco, e non posso scrivere niente qui. Voglio solo dire che mi hai insegnato tantissimo, ma più di tutto che con un sorriso si può superare anche la tragedia di un sintetizzatore che decide di esplodere proprio durante la tua sintesi.

Paola sei la prima persona da cui ho avuto la fortuna di imparare a stare in laboratorio, io che sono sostanzialmente un topo da computer. Sono felice che tu sia riuscita a trovare la tua strada perché te lo meriti, e spero un giorno di imparare anche a volare da te.

Vorrei essere stato d'aiuto a Claudia almeno la metà di quanto non sia stata d'aiuto per me. Dietro la tua sensibilità si nasconde una persona d'oro che mi ha sempre aiutato a fare i conti con la realtà. Se tornassi indietro, rifarei sempre la scelta di passare con te questa avventura.

Se c'è una persona che ha trasceso il rapporto tra colleghi, questa è sicuramente Rosita. Sei stata un'amica fidata oltre che una collega competente e disponibile. Il tuo starmi a fianco è stato di un'inenarrabile importanza in ogni momento, da quelli felici a quelli meno felici, da quelli seri a quelli dove la pazzia ha preso completamente il sopravvento. Tutte le risate, le lacrime, gli strumenti dotati di vita propria, e persino i vicendevoli aiuti Zoteriani, non avrei potuto sperare di condividerli con una persona più bella di te.

Come non menzionare poi Corinne, Giorgio e Manuel. Non vi metto insieme perché vi considero meno importanti, ma proprio per dire a tutti e tre che se anche abbiamo condiviso meno momenti

insieme, ciascuno di voi avrà sempre un posto speciale solo per lui/lei nel mio cuore. Siete stati grandi!

Fabrizia, sai che non ti considererò mai meno importante dei miei illustri colleghi scienziati, anche se sono un fissato della scienza, e parlerei solo di questo dalla mattina alla sera (scusa!). Ti ringrazio di avermi spesso aiutato con parole di conforto nei momenti più difficili, ma soprattutto di avermi sempre fatto parlare di "altro".

Voglio ringraziare tutti i miei colleghi del XXVI ciclo, insostituibili compagni di viaggio e il coordinatore della scuola di dottorato Luigi Paduano per averci saputo motivare a fare un grande lavoro.

In particolare voglio ringraziare Matteo; in questi anni di crescita formativa insieme, non avrei mai potuto sperare di avere una persona più capace e preparata a fianco. I discorsi con te sono stati sempre uno stimolo immenso, stimolando sempre la curiosità. Sono felice di poter "atterrare anche questo morto" insieme a te.

Ora voglio uscire dall'università per ringraziare chi mi è stato a fianco da sempre, al di là del mio lavoro. Soprattutto in questo ultimo periodo mi sono alienato, ma loro sono stati sempre lí quando ho avuto bisogno.

Innanzi tutto voglio ringraziare i miei amici, Alessandro, Davide, Federico, Giulia, Marcella, Mario e Nicola, insostituibili compagni di vita. Stiamo crescendo insieme e non vi voglio perdere mai.

Mariano tu sai quanto vale per me la nostra amicizia, è e sarà sempre un punto fisso della nostra vita. Siamo tutti e due devoti al complicarci la vita. Si dice che se stai con lo zoppo impari a zoppicare, io dico che due zoppi racimolano due gambe buone.

Le mie fantastiche *girls*, Cinzia, Fortuna e Valeria ci sono sempre state e quando dico sempre, intendo al di fuori dello spazio e del tempo. Se in questi anni sono riuscito con immensi sforzi ad imparare a percepire me stesso è soprattutto grazie a voi.

Ringrazio Fra Angelo. Ci ha seguiti come figli, e ci hai annunciato l'Amore di Dio. La Grazia che è sgorgata attraverso di te è un miracolo, e come tale si è moltiplicato nella vita di tutti noi. Grazie, tutto quello che fai non è mai scontato per me.

Mio cugino Bruno e mio fratello Claudio sono parte integrante della mia personalità. Nel bene e nel male ci siamo forgiati l'un l'altro, e li ringrazio perché questa tesi la devo a loro per tanti motivi. Siete i miei fratelli, il mio sangue. Nonna sarebbe fiera di noi.

Mamma e papà per me sono stati le fondamenta senza le quali nulla di tutto questo sarebbe stato possibile. Il loro supportarmi e sopportarmi mi ha rasserenato soprattutto in questo periodo di stress su tutti i fronti.

Marinella è la mia musa e la mia più grande ispirazione. Il tuo amore mi ha mosso ad amare me stesso, a volermi bene anche in questo periodo di dubbi, di difficoltà. Mi sei stata sempre a fianco ed è solo grazie a te se ora posso scrivere queste righe. Vorrei poter sintetizzare con parole l'amore che mi hai dato, quello che mi ha mosso a stare ben saldo in me stesso, solo perchè sapevo dov'ero, ero al tuo fianco. Da quel giorno, qualcosa dentro di me si è spezzato, e tu stai avendo la pazienza di raccogliere i cocci con me. Dovrei scrivere mille parole per te, ma non vorresti, e so che mi stai già odiando per

quel poco che scrivo. Se vorrai, proverò a parlare sempre il tuo linguaggio.

L'ultimo ringraziamento va a te mio Signore, mi hai protetto e coccolato in modi che ancora oggi non riesco a capacitarmene. Sarà sempre nella mia natura stupirmi di fronte a tutto questo infinito amore, questo lo sai. Spero che rendendoti grazie giorno per giorno, possa imparare anch'io a smettere di giudicarmi, come facesti con me quella notte.